EXPLORING HIGH SCHOOL STUDENTS' SCIENCE IDENTITIES, MOTIVATION IN SCIENCE AND ENVIRONMENTAL ATTITUDES

By

ROSA AGHEKYAN

A Dissertation submitted to the

Graduate School-New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Education

written under the direction of

and approved by

Angela O'Donnell

____

____

____

____

____

New Brunswick, New Jersey

May, 2017

ABSTRACT OF THE DISSERTATION

Exploring High School Students' Science Identities, Motivation in Science and Environmental Attitudes

by ROSA AGHEKYAN

Dissertation Director:

Dr. Angela O'Donnell

ABSTRACT

The quality of public education has become the focus of policymakers, educators and researchers since the launch of Sputnik by the Soviet Union (Bybee & Fuchs, 2006; Rutherford, 1997). In particular, science, technology, engineering and math (STEM) are under great scrutiny (Hill, Corbett, & St Rose, 2010). There are valid concerns that the growing generation may not do well in this globalized world due to lack of skills in STEM. Indeed, research showed that US students lag behind their international peers in science. In addition, most US students lack scientific literacy skills (Grigg et al., 2006), as was apparent in their PISA's scores (Bybee & McCrae, 2006). There is much emphasis on mastering scientific content in order to increase scientific literacy and students' international science test scores. This leaves less room for answering questions like "Why don't students do well in science?" In order to improve science education, the root cause

of this problem should be explored. There could be many valid reasons behind this issue. For example, the complexity of scientific phenomena and their intricately interrelated components makes students perceive science as a difficult subject (Aschbacher et al., 2010; Graesser, Singer & Trabasso, 1994). Similarly, many students found science irrelevant to their daily lives (Kadlec, Friedman, & Ott, 2007). Weak science identities and lack of motivation could be another reason. As research showed, students gradually lose their interest in science as they become adolescents (Osborne, Simon, & Collins, 2003). Likewise, students with strong science identities are more likely to participate and succeed in their science classes than students with weak identities (Gresalfi, 2009). Therefore, it is vital to explore students' science identities and motivation in science in order to understand the root causes of science education problems.

Since students' science identities and motivation play such an important role in their science learning, the present study aimed at investigating high school students' science identities, expectations of success in science, and values of science. Additionally, it looked into students' environmental attitudes since environmental issues relate to science learning, and the planned studies of this research were related to the domain of environmental science learning. In order to achieve the study's goals, a new survey instrument, called SIEVEA, was developed. This instrument was used to collect data so further analyses could be conducted. The data was collected in two various contexts: for a large group of students from multiple school districts and states, and for a smaller group of urban high school students participating in a collaborative online project.

The research was made of three studies. Study 1 encompassed the design, development and validation of an instrument called **S**cience **I**dentities, **E**xpectations of

Success in Science, **V**alues of Science and **E**nvironmental **A**ttitudes (SIEVEA). The developed instrument is a convenient online survey that can be used to measure students' science identities, expectations of success in science, values of science, and environmental attitudes. Study 1 was made of three sub-studies. In Study 1A, the SIEVEA was designed and used to collect data of 1,911 high school students (grades nine to twelve) from 11 school districts in New Jersey, Pennsylvania and Connecticut. The collected data was used to validate the survey as a measurement instrument. The data was analyzed using both descriptive statistics and exploratory factor analysis (EFA). The EFA results provided useful insights into the factor structure of the data and led to the formation of three candidate models: the two-factor, the three-factor and the four-factor models. All three models were evaluated based on their fit to data, their alignment with the research constructs, and their factor loadings. As a result of this evaluation, the three-factor model was selected as the final model. In Study 1B, the three-factor model was evaluated using partial-confirmatory factor analysis (PCFA) and confirmatory factor analysis (CFA). The PCFA used the original data, whereas the CFA used a newly collected data, which contained survey responses of 1,495 high school students from three schools (urban and suburban) in New Jersey and Connecticut. Additionally, Study 1B conducted reliability and validity tests on the model, including tests for convergent and discriminant validities. In addition, the instrument was tested for measurement invariance. As a result of these analyses and tests, the three-factor model's selection as the best model was confirmed. Next study, Study 1C, conducted the Rasch analysis on the SIEVEA survey instrument in order to accomplish the following goals: explore the instrument's psychometric properties and find areas of improvement, conduct additional validation tests on the instrument, and convert the ordinal scores of SIEVEA's data to interval scale in preparation of conducting parametric statistical tests.

Study 2 aimed at measuring and analyzing the three constructs related to student science learning as suggested by the three-factor model: students' science identities and motivation in science, values of science and attitudes toward the environment. In order to do this study, two data sets collected by SIEVEA were combined into a single sample. This study showed that students' science subject preferences influenced their science identities and motivation in science. In addition, significant gender-related differences were discovered in students' science subject preferences, science identities and motivation, and values of science. Males had stronger science identities and motivation in science than females, whereas females ascribed higher value to science. Surprisingly, there were no statistically significant differences between males' and females' environmental attitudes. Lastly, statistically significant differences were found between urban and suburban students' environmental attitudes and science subject preferences.

Finally, Study 3 scrutinized how urban high school students' science identities shifted during their participation in a project based on an online, collaborative learning environment called the River City. There were 8 student participants in this project, which took about two weeks. The project facilitated student learning of scientific inquiry and 21st century skills via a game-based, multi-user virtual environment (MUVE). The study's results suggested that students' science identities were not stagnant, but rather that they could change and evolve. However, since this was a short-duration project and due to measurement errors, these results were not conclusive. An additional, longitudinal research is recommended in order to confirm this study's findings.

Dedication

This dissertation is dedicated to the memory of my beloved parents:

Ernest Aghekyan and Melania Muradyan.

Acknowledgements

I would like to thank my advisor, Dr. Angela O'Donnell, who was a great mentor for me, whose guidance, and faith in me consistently directed my work to the right direction. She always understood me and was next to me whenever I needed encouragement and support. I also want to express my deepest gratitude to my committee members, Dr. Clark Chinn, Dr. Chia-Yi Chiu, Dr. Beth Rubin, and Dr. Anselm Spoerri. My dissertation committee members provided me with valuable feedback, which helped me in revising the design of my study. Likewise, my committee was patient with me when I had a delay in my data collection process.

When I look back and reflect on my time at Rutgers, I realize that I have learned a lot, and have gained great experience and confidence, that are necessary components of success. I feel that my journey at Rutgers started with baby steps, such as with the completion of the pilot study about the questioning techniques, followed by more advanced study, which examined the correlation between the higher-order thinking questioning and scientific modeling. Rutgers gave me the opportunity to familiarize myself with the process of publishing and reviewing, which was an experience where I learned enormously from other experts in my field. Furthermore, I have gained experience with the IRB processes, making cold calls to numerous district superintendents and their assistants across various states, while asking permissions for data collection in their schools. Because of my research confidentiality and anonymity protocol, I will not be able to thank to specific people. However, I will generalize and direct my gratitude to the all superintendents, assistant superintendents, science supervisors and science teachers who allowed and supported me to collect data in their high schools. In addition, I want to thank Rick Noll, the owner of the ActiveWorlds, Inc., who gave me access to the River City and helped in supporting and troubleshooting the application whenever there was a need. Likewise, I am thankful to Dr. Linda Paslov, Dr. Amanda R. Bozack, and Dr. Cindy Kern for helping me in finding leads for data collection in Connecticut.

I also want to take this opportunity and thank my immediate family, my husband Aram, my daughter Lilia and my son Michael, for their love, encouragement and support. Thank you for being so patient with me. I acknowledge that I could have spent more time with my family than I did in these past eight years. However, I am hopeful that my dedication to my graduate work will be rewarding and my family members will never regret for those days that I was not able to spend with them.

And my last words go toward my lovely parents, to whom I am dedicating my dissertation. I have to admit, that I was fortunate to grow up in the family that was able to provide me with everything. My dad was a CEO for more than twenty years in three different companies, and he always modestly contributed his successful career to his education. Likewise, my mom, who was a math teacher, contributed her reputation as a great teacher to the pedagogical institution that she had attended. Sadly, both of my parents are not alive, but I am sure that they would have been so proud of me if they would have witnessed my accomplishments. Despite my parents not being next to me, the lessons that I have learned from them will always be carried with me. My parents were great role models for me, and I have learned from them to be passionate, persistent, kind and sympathetic toward others.

TABLE OF CONTENTS

List of Tables

List of Figures

Chapter 1. Introduction

Skills required of workers in the 21st century differ significantly from those needed in preceding centuries (Bybee & Fuchs, 2006; Hofstein & Lunetta, 2004; Silva, 2009). Indeed, to be successful in today's high-tech, science-driven economy, one must be technology savvy, skillful in data analysis, and scientifically literate. Even though science, technology, engineering and mathematics (STEM) have been playing a prominent role for decades in preparing students to become part of modern workforce, there is an even greater focus on STEM education now than ever before. One important reason behind this scrutiny is the continued lag of U.S. students behind their international peers. As the results of The Program for International Student Assessment (PISA) for 2012 demonstrated, there was almost no change in 15-year-old U.S. students' science scores since 2003 (495 in 2003 vs 497 in 2012). This is especially disheartening, given that their counterparts from other countries showed significant improvements in their science scores (PISA, 2012). Even worse, according to the 2009 NAEP report, "Only 21% of 12th grade students scored at or above the proficient level on the NAEP science assessment" (https://nsf.gov/nsb/sei/edTool/data/highschool-06.html). Additionally, there is a growing disparity between ever-growing numbers of STEM related jobs and the lack of student interest pursuing STEM related careers (http://www.ed.gov/stem). According to the U.S. Department of Education, only 28% of high school freshmen demonstrate interest in STEM careers. Moreover, 57% of these students lose interest by the time they graduate high school (http://www.ed.gov/stem).

What can be done to improve students' attitudes towards science learning and push them towards choosing science careers? Lack of motivation and absence of strong science identity are some of the reasons why students are not doing well in science. Motivation, defined as expectations of success and values, and interest are some of the key factors in initiating and sustaining students' interest in science learning (Eccles et al., 1998; Pintrich, 2003; Renninger, Hidi, & Krapp, 2014). Regrettably, research showed that students' interest toward science gradually diminishes as they become adolescents (Osborne, Simon, & Collins, 2003). For example, this occurs as students move to secondary school (Osborne et al., 2003; Tytler, 2007). Hence, it is important to understand what motivates children to learn science in their early ages, in order to develop adequate strategies in igniting interest toward science in students' older ages, when their science interest declines (Bathgate, Schunn & Correnti, 2013). Additionally, students with strong science identities are more likely to participate and succeed in science and math classes (Gresalfi, 2009; Sfard & Prusak, 2005). Therefore, it is important to understand how students' science identities form and evolve over time. Also, the study looked into students' environmental attitudes because environmental issues relate to science learning, and the planned studies were related to the domains of biology and environmental science.

The theoretical framework of this dissertation is built on the Eccles, Wigfield, & Schiefele (1998) Expectancy-Value motivational theory and on the Carlone and Johnson (2007) model of science identities. The following discussion reviews the notions of motivation, identity, interest and environmental attitudes. Then, it introduces the present study with its rationale and research objectives.

Motivation

Motivation can be defined as the study of why people think and act the way they do. In an academic context, motivation plays an important role in explaining the reasons behind why some students complete the task despite its arduous nature, while others give up on easy tasks (Graham & Weiner, 1996). Motivational theories claim that motivational factors influence cognitive processes that, in turn, affect performance (Graham & Weiner, 1996). Therefore, understanding high school students' science identities and their motivation in science can aid in explaining phenomena such as disliking science or not doing well in science.

Eccles et al. (1998) Expectancy-Value framework is a relatively modern theory of motivation built on theoretical constructs from various theorists such as Tolman (1932), Lewin (1938) and Atkinson (1957, 1964). Tolman's (1932) studies in behavioral psychology showed how expectancies for success operate in various areas. Lewin (1938) examined how the value of a task and its importance are connected. According to this theory, the level of aspiration directs individuals' cognitive decision-making processes about engaging or not engaging in tasks. Moreover, individuals set up goals for themselves based on their past experiences and familiarity of the task. Atkinson (1957, 1964) developed the first, mathematical Expectation-Value Theory, which provided an explanation for achievement behaviors and linked persistence and achievement performance to individual's expectancy and value beliefs. Atkinson considered expectancies and values as task specific and closely related to each other. According to Atkinson, individuals mostly value tasks they consider difficult to complete. In Atkinson's theory of achievement, motivation achievement motives, expectancies for success and incentive values are three factors that define achievement behaviors.

Achievement motives included both success achievement and failure avoidance, whereas expectancies for success were the expected probability for the individual to succeed on a task. Lastly, Atkinson defined the incentive value as attractiveness to the task (Wigfield, Tonks & Klauda, 2009).

Jackeline Eccles and her colleagues expanded Atkinson's theory to the educational field. According to Eccles et al. (1998) Expectancy-Value motivational framework, students' achievement behavior is primarily influenced by two factors: their expectation of success in completing the task and the task's value. Eccles et al. (1998) clearly defined and separated expectancy and task value beliefs and, in addition, scrutinized psychological, social and cultural factors of modern Expectancy-Value theory. In this motivational theory, expectancy and task value are considered the most important predictors for students' achievement behavior including achievement-related choices (Eccles et al., 1983). Expectancies are students' beliefs about their abilities to accomplish certain tasks. Thus, these expectancies reflect students' views on their task competence, expectations for future performance, and self-efficacy (Fredricks & Eccles, 2002). In other words, the expectancy element itself is viewed as a product of competence and self-efficacy (Plante, O'Keefe, & Théorêt, 2013). As Schunk, Pintrich and Meece (2008) defined, "Expectations are people's beliefs and judgments about their capabilities to perform a task successfully" (p. 44) and "Values refer to the beliefs students have about the reasons they might engage in a task" (p. 44).

The other component of this theory, task value, relates to students' beliefs about reasons and incentives for completing the task: do they think that doing the task is valuable for them? It should be noted that task values are considered subjective, "because various individuals assign different values in the same activity; math achievement is valuable to some students but not to others" (Wigfield, Tonks, & Klauda, 2009, p. 57). Consequently, individuals will do tasks they positively value and will avoid tasks they do not value (Eccles et al., 1983; Wigfield & Eccles, 1992). Eccles et al. (1983) further decomposed Subjective Task Value (STV) into four components: interest (enjoyment), attainment value (importance), utility value and cost. STV is related to personal and social identities and identity formation processes (Eccles, 2009; Wigfield & Eccles, 1992). Indeed, adolescents' choices of certain activities are often based on preferences of groups that they relate to, their culture and individual experiences. Therefore, the values of these activities are driven by the same factors that play a big role in identity formation (Eccles, 2009).

Intrinsic value/interest is the enjoyment that one receives from a task (Wigfield, et al., 2009), the pleasure that the individual receives as a result of engaging in an activity (Eccles, 2009) or the subjective interest in an activity as stated by some scholars (Plante et al., 2013). Hulleman, Durik, Schweigert & Harackiewicz (2008) state that finding a certain activity enjoyable (e.g., a lecture or a reading assignment) is a sign of intrinsic value. Attainment value is the personal importance of completing the task (Plantae et al., 2013) and can be established by underlying needs to achieve desired outcomes (McClelland, 1961). Attainment value is the intrinsic importance of the task to individuals (Eccles, 2005) or the importance of being good at something (Eccles & Wigfield, 1995). If students are good at something, they will value that task (Jacobs, Lanza, Osgood, Eccles, & Wigfield, 2002). Eccles et al. (1983) connected attainment value to the individual's self-schema, because tasks make it possible to express different aspects of one's self-schema. This is why, not surprisingly, attainment value encompasses individual's identity matters: how the task relates to the individual's self (Wigfield et al., 2004).

Utility value is how an individual relates current activities to goals, such as future occupation and career (Wigfield & Eccles, 1992). Utility value can also relate to current goals (Plante et al., 2013). Utility value is associated with extrinsic motivation (e.g., an individual does not like math but still takes a math course, since it is a requirement for college admission) but it can be connected to intrinsic motivation too when an individual is truly interested in the task due to personal goals (Wigfield et al., 2004). For this reason, both intrinsic motivation and interest are considered valuable educational outcomes (Durik & Harackiewicz, 2007). Cost value encapsulates many elements, including lost opportunities due to making one choice instead of another, the negative aspects of engaging in a task, like fear of failure (Plante et al., 2013), and the amount of effort devoted toward accomplishing the task (Wigfield & Eccles, 1992). Cost is also considered an essential component of choice making (Eccles et al., 1983) and, interestingly, it is the least researched element of subjective values (Wigfield et al., 2004).

Research showed that students' beliefs and expectations about themselves influence their motivation (Pintrich & Schunk, 2002). Holding certain beliefs about the self plays a major role in individuals' motivation and well-being. For instance, students who have higher expectations, stronger beliefs about themselves and more confidence about their abilities are more motivated and cognitively engaged than their peers who lack these behaviors (Bandura, 1997; Eccles et al., 1998). Past research indicates that although both expectancy beliefs and values are correlated with high achievement, expectancy beliefs contributed more to achievement than did value beliefs (Eccles et al., 1983).

People's judgment regarding their abilities on specific tasks demonstrates their self-efficacy. Some people can be very confident about their skills and accomplishments, while in reality they may be quite moderate. On the other hand, other individuals may doubt their abilities, despite possessing certain capabilities. Very often, human behavior can be predicted by the beliefs they hold. Bandura (1997) has pointed out that the individual's beliefs, rather than the realistic assessment of the individual's expertise, shape personal accomplishments. Not surprisingly, people who hold negative thoughts and fears think that they cannot be successful and avoid arduous tasks. They also most likely will experience lower efficacy than people who think positively about themselves and, as a result, possess high self-efficacy.

It is not only important how individuals perceive themselves, but also how individuals are viewed by others. Harter (1988) observed that when an individual has a perception of others' negative attitude toward him/her, it gets internalized as a low self-opinion. Furthermore, low self-opinion can lead to lower achievement. Harter argued that students with high-perceived academic competence are more likely to be high academic achievers than students with low-perceived competence.

Identity

Identity is one of the most discussed and researched subjects in modern educational research. It used to be a specialized research topic in psychology. Currently, it is employed in various fields, including anthropology, sociology, history, cultural studies and education (Sfard & Prusak, 2005). Academic identities, such as those present in various science domains, have interested educational researchers for some time, as they allow researchers to understand why people make certain educational choices (Tucker-Raymond, Varelas, Pappas, Korzh, & Wentland, 2007).

Many educational researchers tried to define and conceptualize the construct of identity and explain students' learning via identity and its evolution (Moje, Tucker-Raymond et al., 2007). Their studies explored the significance of identity as both research and as an educational construct. Identity research focuses on several key questions: What is identity? What are science identities? How do identities and science identities form? Do identities and science identities shift, evolve or stay stagnant? Are there explicit connections between science identities and science learning?

In addition to discussing the conceptual aspects of identity, the research also looked into its practical applications, including teaching practices that can shape students' science identities and ways in which science identities influence student learning. In this light, other questions need to be answered. How do we know if we have a science identity? How do students perceive scientists and posit themselves as science learners in science classrooms? Does age matter for science identity formation? Is how students view themselves in science learning or how their teachers and peers view them more important? Although the list of questions goes on, it makes sense to stop here and turn to extensive literature to reflect on how the notions of identity and science identities are defined.

Identities can be defined as narratives enacted in time, space, and relationships (Bruner, 1991; Moje et al., 2007). Tucker-Raymond et al. (2007) argued that narratives as enactments of identities are not “pure,” meaning they are not just “discourse” and no “action,” but that they rather include both. The interactions during these enactments are simply subject positionings; therefore, they are referred to as subjectivities. Since the enactments happen in different settings, identities can be referred to as “multimodal” narratives (Tucker-Raymond et al., 2007). The “multimodality” of identities necessitates data collection in various ways: storytelling through words, pictures, and narratives (Moje et al., 2007). Learning and identities are related to each other (Tucker-Raymond et al., 2007). Learning involves more than construction of information and juxtaposing of different pieces of information to paint the whole picture. It is also about people thinking about themselves in relation to the process of learning concepts (Moje et al., 2007).

Identity is a multi-faceted notion in educational research, with various definitions existing across a wide range of conceptual frameworks and constructs. Indeed, Gee (2000) defined identity as being seen as a certain “kind of person.” This view of identity focuses on the individual’s performance in society and how one is recognized by himself/herself and by others. According to Gee, individuals can have multiple identities in addition to their core identities. In a science context, this translates into the individual’s science identity defined as that of a “science person.”

Carlone and Johnson (2007) developed a model of science identity by projecting students’ science identities onto three overlapping dimensions: science competence, science performance and recognition as a science person. According to this model, in order to be considered an individual with a strong science identity, a student should rate himself/herself highly and be rated highly by others in all three dimensions.

Students' science identities play an important role in their participation in science and math classes (Gresalfi, 2009; Sfard & Prusak, 2005). Science identity is an important tool for understanding student science learning, due to the fact that science identities form through practice, which requires knowledge, skills, social context and domain-specific ways of thinking (Barton et al., 2013). By apprehending students' science identities, science educators can get a better understanding of their students' personalities and aspirations. Through this, they can develop more meaningful and beneficial instruction (Kane, 2016). Science identity can be used as an analytical tool for investigating science learning; more so, identity serves as "the missing link" between learning and sociocultural context (Sfard & Prusak, 2005).

Since the approaches to identity and science identity in current educational research are numerous and vary by researchers, it is helpful to capture and organize them in a table (see Table 1.1).

Table 1.1
Summary of Different Notions of Identity/Science Identity

Notion of Identity	Study
"…consider identity to be a social semiotic sign that, like any sign, has both an inner life in people's minds and an outer representation"	Tucker-Raymond et al. (2007, p. 560)
"Learning is, in this purview, more basically, a process of coming to be, of forging identities in activity in the world."	Lave (1992, p. 3)
"…the notion of identity proves helpful in dealing with issues of power and of personal and collective responsibilities for individual lives. In particular, identity features prominently whenever one addresses the question of how collective discourses shape personal worlds and how individual voices combine into the voice of a community."	Sfard & Prusak (2005, p. 4)
"Individual identity is not necessarily either single or stable. A person can be a part of or aspire to many different communities simultaneously."	Brickhouse, Lowery, & Schultz (2000, p. 443)

“...identity, knowledge, and social membership entail one another.”	Lave & Wenger (1991, p. 53)
”children are born into a particular society and become competent, functioning individuals with particular social identities to the extent that they re-construct for themselves the social representations of the significant groups in their society.”	Lloyd & Duveen (1992, p. 27)
“Being recognized as a *certain ‘kind of person*,’ in a given context, is what I mean by ‘identity’”	Gee (2000, p. 99)
“Identity is an “internalized positional designation”	Stryker (1980, p. 60)
“The experience of identity in practice is a way of being in the world.”	Wenger (1998, p. 151)
“identity is…individual agency as well as societal structures that constrain individual possibilities.”	Brickhouse (2000, p. 286)
“identity as an analytic lens involves new ways of viewing the process of learning, as the socialization of students into the norms and discourse practices of science”	Carlone & Johnson (2007, p 1189)
“People construct identities as they communicate (in words and actions) the ways in which they perceive themselves (and the ways they believe others perceive them) in relationship to particular individuals, groups, ideas, activities, institutions, and so on.”	Kane (2012, p. 458)
“Identity can be considered an enactment of self made within particular activiites and relationhsips…”	Moje (2004, p. 16)
“Learning science is thus manifested through the transformation of ‘identity-in-practice’ in the science classroom”	Tan & Barton (2008, p. 6)

Psychological theorists stated that there is a linkage between motivation and identity (Eccles, 2009; La Guardia, 2009; Renninger, 2009). For instance, Eccles (2009) made connections between expectancy-value theory and personal as well as social/collective identities, arguing further that personal identities, such as a unique self, social/collective identities and social groups, are the reflection of individual’s choices and behavioral engagement (task value and competence). Likewise, Le Guardia (2009) drew a link between three essential components of self-determination theory and the formation of healthy identities by demonstrating that important psychological needs of autonomy, competence and relatedness internalize and conceptualize the conditions that foster the individual's identity.

Interest

Traditionally, the role of interest and its impact on learning were considered important topics in psychology and in education as well (Renninger, Hidi, & Krapp, 2014). The central challenge in defining interest is its overlapping characteristics with other similar constructs like motivation and engagement. Indeed, almost all definitions of interest imply some kind of motivational component along with engagement with a task/activity that is the target of interest. Certainly, motivation and interest are mutually related, because interest can be thought of as a precursor to intrinsic motivation and, likewise, mastery goal adoption can lead to interest generation. Interest and engagement also seem to be related since students' interest in a given task should lead to a high level of engagement in that task; however, their connection is not conclusive (Ainley, 2004).

Notwithstanding all similarities among interest, motivation and engagement, interest is a distinct concept, and some researchers tried to differentiate it by pointing to its relation to specificity (Krapp et al., 1992; Renninger, 2000). According to these researchers, interest is object, content, or domain specific, meaning its outcomes are connected to specific relations. This greatly compares with motivation, which is applicable to more generalized behavior. Yet, very often, "interest" and "motivation" are used interchangeably in colloquial language, despite their differences (Schiefele, 2009). Schiefele argued further that motivation is influenced by motives and goal orientations and embodies an individual's desire to accomplish the task, whereas interest "represents a possible antecedent of motivation" (p. 197).

Even though interest is a well-understood term in everyday language, its definition varies greatly among different researchers. Depending on the theoretical perspective of a researcher, interest is assigned a differing meaning. Table 1.2 contains some examples of how interest is defined by various researchers.

Table 1.2
Various Definitions of Interest

Definition	Study
"a motivational variable" referring to "the psychological state of engaging or the predisposition to reengage with particular classes of objects, events, or ideas over time" (p.112)	Hidi & Renninger (2006)
"interest is conceptualize as the core affect of the self – the affect that relates one's self to activities that provide the type of novelty, challenge, or aesthetic appeal that one desired at the time" (p.45)	Deci (1992)
"a content-specific motivational characteristic" (p.299)	Schiefele (1992)
"liking and willful engagement in a cognitive activity" (p.23)	Schraw & Lehman (2001)
"a motivation to engage with a topic (e.g., dinosaurs) or an activity (e.g., photography)" (p.26)	Edelson & Joseph (2004)
"Genuine interest is the accompaniment of the identification, through action, of the self with some object or idea, because of the necessity of that object or idea for the maintenance of a self-initiated activity" (p.14)	Dewey (1913)

Ryan and Deci (2000) classified interest and enjoyment as subcategories of intrinsic motivation. Likewise, Silvia (2001) differentiated the concepts of "interest" and "interests" by clarifying "interest" as emotions and "interests" as certain motives held by people for engaging in certain activities. Research distinguishes personal and situational interest. Pintrich (2003) considered personal interest to be a stable variable, which represents the individual's enduring attraction toward and enjoyment during a particular activity. Pintrich further distinguished personal interest from curiosity by elaborating that curiosity is "assumed to be a personal characteristic of the person, but it is more diffusely directed toward many different activities (e.g., a student who is curious about many different topics)" (Pintrich, 2003, p. 674). Furthermore, personal interest, also known as individual interest, can be further conceptualized as dispositional and as actualized. Dispositional interest lasts for a long period of time and influences learning in a variety of situations, whereas actualized interest "can be said to 'show itself' in particular psychological states, such as focused, prolonged, relatively effortless attention" (Renninger, Hidi & Krapp, 2014, p. 7). Misconceptions exist in research that if interest is absent, then it cannot be developed. This is the result of ignoring the whole picture, such as how and why interest develops (Hidi & Renninger, 2006).

Situational interest, on the other hand, gets invoked as a result of an interesting task or activity (Pintrich & Schunk, 2002). Researchers classified two main types of situational interest: triggered and maintained (Hidi & Baird, 1986; Mitchell, 1993). Triggered situational interest, as its name suggests, is triggered by external factors and is usually short-term, Maintained situational interest is invoked when the learner finds the task meaningful and has the ability to "hold" the interest (Schiefele, 2009). Because of its capability of "holding" interest, the maintained interest has the potential of being transferred to long-term individual interest (Hidi & Renninger, 2006). This fits well within Hidi and Renninger's (2006) "The Four Phase Interest" model, which maintains that the learner's interest can gradually deepen by evolving through the following phases: triggered situational interest, maintained situational interest, emerging individual interest and well-developed individual interest. As motivational constructs, both situational interest and curiosity are heavily influenced by the environment (Renninger et al., 2014). Another type of situational interest, called text-based interest, can occur only once or repetitively as a result of surprise, seduction and other amazement provoking sentences in text (Hidi & Baird, 1986). Likewise, Silvia (2005) suggested situational interest be conceptualized as an emotion, as interest encapsulates behavior, facial expressions, fascination, and engagement.

Environmental Attitudes

Various institutions emphasize the importance of environmental literacy and educating environmentally literate citizens (Jowett, Harraway, Lovelock, Skeaff, Slooten, Strack & Shephard, 2014). Indeed, the goal of environmental education is to increase students' environmental literacy and raise their awareness of environmental problems. This is especially because people's habits and choices of lifestyle in the 21st century may result in the destruction of the environment and in the diminishment of resources (Erdogan, Bahar & Usak, 2012). It is important for students to familiarize themselves with their local environment first, before making judgments regarding global environmental issues, such as global warming and water/air pollution (Sobel, 1996).

Environmental concern is a psychological factor that refers to the extent to which the individual expresses concerns about the environmental problems and acknowledges their impact (Dunlap, Van Liere, Mertig, & Jones, 2000). Environmental scholars highlighted the link between environmental science and behavioral science by exploring how human behavior contributed to widely spread environmental problems, such as global warming, water/soil/air pollution and biodiversity loss (Gardner & Stern, 2002).

Previous studies showed that students who held strong environmental attitudes also demonstrated positive behavior toward education and society (Goldman, Yavetz & Peer, 2006). However, not enough research was done to evaluate how culture and individual values influence the formation of the link between environmental attitudes and behavior (Eom, Kim, Sherman, & Ishii, 2016).

Research claims that strong attitudes are more stable over time than weak ones (Krosnick & Petty, 1995) and are more common among older rather than younger individuals (Alwin, Cohen & Newcomb, 1991). Coyle's (2005) report regarding environmental literacy collected from ten years of surveys not only showed that most of the Americans were environmentally illiterate, but also concluded that even small changes in human behavior can have a huge impact on environment. One may ask why there should be so much attention toward the environmental education. The answer to this question is that it will not only increase students' environmental awareness, but in general it will leave a positive impact on students (Aldous, 2010). Like Project 2061's (AAAS, 1989) emphasis on the growth of scientific literacy, several studies highlight the importance of cultivating environmental literacy, where individuals gain better perception regarding the interaction between the humans and their surrounding environment (Osbaldiston, 2004).

The relationship between attitude and behavior was the center of psychological and sociological research for many years. Despite this, there was not enough research conducted in the environmental education field to examine the relationship between environmental attitudes and behavior (Eilam & Trop, 2012). This is regrettable, as fostering positive environmental attitudes in students is one of the primary objectives of environmental education. In addition, environmental attitudes are closely related to environmental awareness. Indeed, the Aminrad, Zakariya, Hadi, & Sakari (2013) study demonstrated a strongly positive relationship between environmental awareness and environmental attitude (correlation r = .99). This study was a survey conducted to measure relationships between environmental attitude, environmental knowledge and environmental awareness (AKA) of secondary students in Malaysia. The researchers attributed students' positive environmental attitude, environmental knowledge and environmental awareness to their families, teachers and media.

Likewise, an increase in students' environmental literacy leads to greater environmental awareness with respect to modern environmental issues, like climate change, pollution, and habitant and biodiversity loss (Erdogan, Bahar & Usak, 2012). For example, Cornell Professor David Pimentel's research finding that 40% of deaths worldwide are related to water, air and soil pollution (Lang, 2007) is a startling topic that ignites interest and curiosity in students to learn more about their own environment and pose reasonable solutions toward pollution prevention. Interestingly, past research did not focus on gender differences in environmental attitudes and knowledge (Gough, 1998), although studies that examined the gender effect anticipated that females would demonstrate better environmental attitudes than males. Even though no empirical data was provided in support of this claim, some researchers presumed this expected difference because females are perceived to be more altruistic and have greater socialization needs (Gilligan, 1982).

As Bradley, Waliczek, and Zajicek (1999) study's results indicated, there was a gain in environmental knowledge, as well as a positive attitude change toward the environment, after high school students were introduced to a short environmental science course. Jowett et al. (2014) used the revised New Ecological Paradigm’ (NEP) scale to monitor how students’ environmental concerns change. These scholars tried to understand the meaning of environmental literacy and the criteria that would show the progress of environmental education. Interestingly, simple statistical analysis did not find significant differences between the mean NEP scores of the first year and second year students. However, when the multinomial-regression model was used, the authors noticed an overall improvement in students’ beliefs regarding environmental concerns. For example, students that were neutral toward environmental problems in their first year in college were re-categorized as “green” in their second year.

The Present Study: Exploring High School Students’ Science Identities, Motivation in Science and Environmental Attitudes

The primary goal of the present study was to investigate high school students’ science identities, expectations of success in science, values of science and environmental attitudes. As the prior discussion explained, students’ science identities, expectations of success in science, and values of science play an important role in their science learning. Therefore, it was important to explore these constructs within a large and diverse sample of students. In order to achieve this goal, it was critical to have an instrument that could validly and reliably measure these constructs for a large group of students. Once developed and validated, the instrument could have been used to measure these constructs in various contexts: either for a large group of students from different school districts and states or for a smaller set of students participating in a collaborative online project.

Study 1 of this research was made of three sub-studies which were labeled accordingly as Study 1A, Study 1B and Study 1C. In Study 1A, a simple instrument for measuring students' science identities, expectations of success in science, values of science and environmental attitudes was designed and thoroughly validated. The instrument was an online survey called SIEVEA. Additionally, the survey collected students' gender data and their favorite science subjects. The survey's design was based on prior studies related to the aforementioned constructs and instruments used for measuring them. First, the survey was piloted and modified based on pilot's results. Then, it was administered to 1,911 high school students from 11 school districts in New Jersey, Pennsylvania and Connecticut. The collected large sample was used to conduct exploratory factor analysis (EFA) in order to validate the factor structure and link it to the research constructs. During the EFA, three candidate models were considered: the two-factor, the three-factor and the four-factor models. After analyzing models' factor loadings, their alignment with the research constructs and their fit to data, the three-factor model was selected as the final model. Next, the three-factor model was further scrutinized using partial-confirmatory factor analysis (PCFA) and confirmatory factor analysis (CFA) and an additional data set of 1,495 high school students' responses, which was independently collected. These analyses confirmed the appropriateness of the three-factor model's selection made during the EFA. Conducting the PCFA and CFA was the primary objective of Study 1B. In addition, Study 1B carried out extensive and comprehensive validity and reliability tests of the survey. Furthermore, in Study 1C the Rasch analysis were conducted to perform in-depth psychometric analyses of the survey and convert the ordinal scores of the survey's data to interval scale in preparation for conducting parametric statistical tests. The primary research methods of studies 1A, 1B and 1C were quantitative methods using various statistical techniques.

Next, in Study 2 the survey data was used to explore students' science identities and motivation in science, their attitudes toward the environment, and their values of science. These constructs were analyzed along with students' demographic data: students' gender and school type (urban vs suburban). The study investigated whether there were any gender or school type related differences in high school students' science identities and motivation in science, values of science and attitudes toward the environment. Study 2 also looked into students' science subject preferences. It examined how students' science subject preferences vary by gender and school type. Additionally, the study scrutinized how students' science subject preferences influence their science identities and motivation in science, their values of science and their environmental attitudes. This study again used quantitative research methods like ANOVA and a Chi-square test.

The last study of this research, Study 3, examined how urban students' science identities change and evolve in an online, collaborative learning environment called the River City. The River City is a multi-user virtual environment (MUVE) which facilitates student learning of scientific inquiry and 21st century skills via a game-based, interactive user interface. In this study, the SIEVEA survey was one of instruments used for measuring students' science identities. The results of this study indicated that students' science identities were not stagnant, but rather that they evolved. However, due to the short duration of this project and measurement errors, its results were not conclusive. The primary research methods of Study 3 were qualitative methods.

References

Ainley, M. D. (2004). *Some perspectives on interest in learning and classroom interaction.* Retrieved from http://www.aare.edu.au/98pap/ain98054.htm

Aldous, D. E. (2010). The people-plant-place paradigm. In J. Janick, G. R. Dixon, & L. Rallo (Eds.), *XXVIII international horticultural congress on science and horticulture for people (IHC2010): Colloquia and overview* (pp. 171-178). Retrieved from http://www.actahort.org/books/916/

Alwin, D. F., Cohen, R. L., & Newcomb, T. M. (1991). *Political attitudes over the life span: The Bennington women after fifty years.* Madison, WI: University of Wisconsin Press.

American Association for the Advancement of Science. (1989). *Science for all Americans*. New York, NY: Oxford University Press. Also available at http://www.project2061.org/publications/sfaa/online/sfaatoc.htm

Aminrad, Z., Zakariya, S. Z. B. S., Hadi, A. S., & Sakari, M. (2013). Relationship between awareness, knowledge and attitudes towards environmental education among secondary school students in Malaysia. *World Applied Sciences Journal, 22*(9), 1326-1333.

Aschbacher, P. R., Li, E., & Roth, E. J. (2010). Is science me? High school students' identities, participation and aspirations in science, engineering, and medicine. *Journal of Research in Science Teaching, 47*(5), 564-582.

Atkinson, J. W. (1957). Motivational determinants of risk-taking behavior. *Psychological Review, 64*, 359-372.

Atkinson, J. W. (1964). *An introduction to motivation*. Princeton, NJ: Van Nostrand.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Macmillan.

Barton, A. C., Kang, H., Tan, E., O'Neill, T. B., Bautista-Guerra, J., & Brecklin, C. (2013). Crafting a future in science tracing middle school girls' identity work over time and space. *American Educational Research Journal*, *50*(1), 37-75.

Bathgate, M. E., Schunn, C. D., & Correnti, R. (2014). Children's motivation toward science across contexts, manner of interaction, and topic. *Science Education, 98*(2), 189-215.

Bradley, J. C., Waliczek, T. M., & Zajicek, J. M. (1999). Relationship between environmental knowledge and environmental attitude of high school students. *The Journal of Environmental Education, 30*(3), 17-21.

Brickhouse, N.W. (2000). Embodying science: A feminist perspective on learning. *Journal of Research in Science Teaching, 38*, 282–295.

Brickhouse, N., Lowery, P., & Schultz, K. (2000). What kind of a girl does science? The construction of school science identities. *Journal of Research in Science Teaching, 37*, 441–458.

Bruner, J. (1991). The narrative construction of reality. *Critical Inquiry, 18,* 1-21.

Bybee, R.W., & Fuchs, B. (2006). Preparing the 21st century workforce: A new reform in science and technology education. *Journal of Research in Science Teaching, 43*(4), 349-352.

Bybee, R., & McCrae, B. (2011). Scientific literacy and student attitudes: Perspectives from PISA 2006 science. *International Journal of Science Education*, *33*(1), 7–26.

Carlone, H. B., & Johnson, A. (2007). Understanding the science experiences of successful women of color: Science identity as an analytic lens. *Journal of Research in Science Teaching*, *44*(8), 1187-1218.

Coyle, K. (2005). *Environmental literacy in America: What ten years of NEETF/Roper research and related studies say about environmental literacy in the U.S.*. Washington, DC: The National Environmental Education & Training Foundation. Retrieved from http://www.neefusa.org/pdf/ELR2005.pdf

Deci, E. L. (1992). The relation of interest to the motivation of behavior: A self-determination theory perspective. In K. A. Renninger, S. Hidi, & A. Krapp (Eds.), *The role of interest in learning and development* (pp. 43-70). Hillsdale, NJ: Erlbaum.

Dewey, J. (1913). *Interest and effort in education.* Houghton Mifflin.

Dunlap, R. E., Van Liere, K. D., Mertig, A. G., & Jones, R. E. (2000). New trends in measuring environmental attitudes: measuring endorsement of the new ecological paradigm: A revised NEP scale. *Journal of Social Issues*, *56*(3), 425-442.

Durik, A. M., & Harackiewicz, J. M. (2007). Different strokes for different folks: How individual interest moderates the effects of situational factors on task interest. *Journal of Educational Psychology*, *99*(3), 597.

Eccles, J. S. (2005). Subjective task value and the Eccles et al. model of achievement-related choices. In A. J. Elliot & C. S. Dweck (Eds.), *Handbook of competence and motivation* (pp. 105-121). New York, NY: Guildford.

Eccles, J. S. (2009). Who am I and what am I going to do with my life? Personal and collective identities as motivators of action. *Educational Psychologist, 44,* 78–89.

Eccles, J., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J. L., & Midgley, C. (1983). Expectancies, values, and academic behaviors. In J. T. Spence (Ed.), *Achievement and Achievement motives: Psychological and sociological approaches* (pp. 75-146). San Francisco, CA: W. H. Freeman.

Eccles, J. S., & Wigfield, A. (1995). In the mind of the actor: The structure of adolescents' achievement task values and expectancy-related beliefs. *Personality and Social Psychology Bulletin, 21*(3), 215-225.

Eccles, J. S., Wigfield, A., & Schiefele, U. (1998). Motivation to succeed. In W. Damon (Series Ed.) & N. Eisenberg (Vol. Ed.), *Handbook of Child Psychology: Vol. 3. Social, emotional and personality development* (5th ed., pp. 1017–1094). New York, NY: Wiley.

Edelson, D. C., & Joseph, D. M. (2004, June). The interest-driven learning design framework: motivating learning through usefulness. In *Proceedings of the 6th international conference on Learning sciences* (pp. 166-173). International Society of the Learning Sciences.

Eilam, E., & Trop, T. (2012). Environmental attitudes and environmental behavior—Which is the horse and which is the cart? *Sustainability*, *4*(9), 2210-2246.

Eom, K., Kim, H. S., Sherman, D. K., & Ishii, K. (2016). Cultural variability in the link between environmental concern and support for environmental action. *Psychological Science, 27*(10), 1331-1339.

Erdogan, M., Bahar, M., & Usak, M. (2012). Environmental education in high school 9th-12th biology course curricula started to be implemented in 2007. *Educational Sciences: Theory & Practice. 12*(3), 2230-2235.

Fredricks, J. A., & Eccles, J. S. (2002). Children's competence and value beliefs from childhood through adolescence: growth trajectories in two male-sex-typed domains. *Developmental Psychology, 38*(4), 519-533.

Gardner, G. T., & Stern, P. C. (2002). *Environmental problems and human behavior* (2nd ed.). Boston, MA: Pearson.

Gee, J. P. (2000). Identity as an analytic lens for research in education. *Review of Research in Education*, *25,* 99-125.

Gilligan, C. (1982). *In a different voice.* Harvard University Press.

Goldman, D., Yavetz, B., & Peer, S. (2006). Environmental literacy in teacher training in Israel: Environmental behavior of new students. *Journal of Environmental Education, 38*(1), 3-22. http://dx.doi.org/10.3200/JOEE.38.1.3-22

Gough, A. (1998). Researching environmental education in teacher education: Initiating and sustaining student interest. *International Research in Geographical and Environmental Education, 7*(3), 260-264.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review, 101,* 371–395. doi:10.1037/0033-295x.101.3.371

Graham, S., & Weiner, B. (1996). Theories and principles of motivation. *Handbook of Educational Psychology, 4*, 63-84.

Gresalfi, M. S. (2009). Taking up opportunities to learn: Constructing dispositions in mathematics classrooms. *The Journal of the Learning Sciences*, *18*(3), 327-369.

Grigg, W. S., Lauko, M. A., & Brockway, D. M. (2006). *The nation's report card: Science 2005* (NCES 2006–466). Washington, DC: U.S. Government Printing Office.

Harter, S. (1988). Developmental processes in the construction of the self. In T. D. Yawkey & T. E. Johnson (Eds.), *Integrative processes and socialization: Early to middle childhood* (pp. 45-78). Hillsdale, NJ: Erlbaum.

Hidi, S., & Baird, W. (1986). Interestingness - A neglected variable in discourse processing. *Cognitive Science, 10*(2), 179-194.

Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist*, *41*(2), 111-127.

Hill, C., Corbett, C., & St Rose, A. (2010). *Why so few? Women in science, technology, engineering, and mathematics.* American Association of University Women. 1111 Sixteenth Street NW, Washington, DC 20036.

Hofstein, A., & Lunetta, V. N. (2004). The laboratory in science education: Foundations for the 21st century, *Science Education, 88*, 28-54.

Hulleman, C. S., Durik, A. M., Schweigert, S., & Harackiewicz, J. M. (2008). Task values, achievement goals, and interest: An integrative analysis. *Journal of Educational Psychology, 100*, 398-416.

Jacobs, J. E., Lanza, S., Osgood, D. W., Eccles, J. S., & Wigfield, A. (2002). Changes in children's self-competence and values: Gender and domain differences across grades one through twelve. *Child Development*, *73*(2), 509-527.

Jowett, T., Harraway, J., Lovelock, B., Skeaff, S., Slooten, L., Strack, M., & Shephard, K. (2014). Multinomial-regression modeling of the environmental attitudes of higher education students based on the revised new ecological paradigm scale. *The Journal of Environmental Education*, *45*(1), 1-15.

Kadlec, A., Friedman, W., & Ott, A. (2007). *Important, but not for me: Kansas and Missouri students and parents talk about math, science, and technology education.* New York: Public Agenda. Retrieved from http://files.eric.ed.gov/fulltext/ED498649.pdf.

Kane, J. M. (2012). Young African American children constructing academic and disciplinary identities in an urban science classroom. *Science Education*, *96*(3), 457-487.

Kane, J. M. (2016). Young African American boys narrating identities in science. *Journal of Research in Science Teaching*, *53*(1), 95-118.

Krapp, A., Hidi, S., & Renninger, K. A. (1992). Interest, learning and development. In K. A. Renninger, S. Hidi, & A. Krapp (Eds.), *The role of interest in learning and development* (pp. 3-25). Hillsdale, NJ: Erlbaum.

Krosnick, J. A., & Petty, R. E. (1995). Attitude strength: An overview. In R. E. Petty & J. A. Krosnick (Eds.), *Attitude strength: Antecedents and consequences* (Vol. 4, pp. 1–24). Hillsdale, NJ: Erlbaum.

La Guardia, J. G. (2009). Developing who I am: A self-determination theory approach to the establishment of healthy identities. *Educational Psychologist*, *44*(2), 90-104.

Lang, S. (2007). *Water, air and soil pollution causes 40 percent of deaths worldwide, Cornell research survey finds.* Retrieved from http://www.news.cornell.edu/stories/aug07/morediseases.sl.html.

Lave, J. (1992, April). *Learning as participation in communities of practice*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. New York: Cambridge University Press.

Lewin, K. (1938). *The conceptual representation and the measurement of psychological forces*. Durham, NC: Duke University Press.

Lloyd, B., & Duveen, G. (1992). *Gender identities and education: The impact of starting school*. New York: St. Martin's Press.

McClelland, D. C. (1961). *The achieving society*. Princeton, NJ: Van Nostrand.

Mitchell, M. (1993). Situational interest: Its multifaceted structure in the secondary school mathematics classroom. *Journal of Educational Psychology, 85*(3), 424.

Moje, E. B. (2004). Powerful spaces: Tracing the out-of-school literacy spaces of Latino/a youth. In K. Leander & M. Sheehy (Eds.), *Spatializing literacy research and practice* (pp. 15-38). New York, NY: Peter Lang.

Moje, E. B., Tucker-Raymond, E., Varelas, M., & Pappas, C. C. (2007). FORUM: Giving oneself over to science: Exploring the roles of subjectivities and identities in learning science. *Cultural Studies of Science Education*, *1*(3), 593-601.

Osbaldiston, R. (2004). Meta-analysis of the responsible environmental behavior literature (Doctoral dissertation, University of Missouri—Columbia, 2004). *Dissertation Abstracts International*, *65*(8), 4340B.

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.

Plante, I., O'Keefe, P. A., & Théorêt, M. (2013). The relation between achievement goal and expectancy-value theories in predicting achievement-related outcomes: A test of four theoretical conceptions. *Motivation and Emotion*, *37*(1), 65-78.

Pintrich, P. R. (2003). A motivational science perspective on the role of student motivation in learning and teaching contexts. *Journal of Educational Psychology*, *95*(4), 667-686.

Pintrich, P. R., & Schunk, D. H. (2002). *Motivation in education: Theory, research, and applications*. Englewood Cliffs, NJ: Prentice Hall.

PISA. (2012). *Results in Focus: What 15-year-olds know and what they can do with what they know.* Retrieved from http://www.oecd.org/pisa/keyfindings/pisa-2012-results-overview.pdf

Renninger, K. A. (2000). Individual interest and its implications for understanding intrinsic motivation. In C. Sansone & J. M. Harackiewicz (Eds.), *Intrinsic and extrinsic motivation: The search for optimal motivation and performance* (pp. 375–407). New York: Academic.

Renninger, K. A. (2009). Interest and identity development in instruction: An inductive model. *Educational Psychologist, 44*(2), 105-118.

Renninger, A., Hidi, S., & Krapp, A. (2014). *The role of interest in learning and development.* New York, NY: Psychology Press.

Rutherford, F. J. (1997, October). Sputnik and science education. In symposium, *Reflecting on Sputnik: Linking the past, present, and future of educational reform*. Washington, DC.

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology, 25*(1), 54-67.

Schiefele, U. (1992). Topic interest and levels of text comprehension. In K. A. Renninger, S. Hidi, & A. Krapp (Eds.), *The role of interest in learning and development* (pp. 151–183). Hillsdale, NJ: Erlbaum.

Schiefele, U. (2009). Situational and individual interest. In K. R. Wentzel & A. Wigfield (Eds.), *Handbook of motivation at school* (pp. 197-222). New York, NY: Taylor Francis.

Schraw, G., & Lehman, S. (2001). Situational interest: A review of the literature and directions for future research. *Educational Psychology Review*, *13*(1), 23-52.

Schunk, D. H., Pintrich, P. R., & Meece, J. L. (2008). *Motivation in education: Theory, research, and applications.* New Jersey: Pearson.

Sfard, A., & Prusak, A. (2005). Telling identities: In search of an analytic tool for investigating learning as a culturally shaped activity. *Educational Researcher*, *34*(4), 14-22.

Silva, E. (2009). Measuring skills for 21st century learning. *The Phi Delta Kappan, 90*(9), 630-634.

Silvia, P. J. (2001). Interest and interests: The psychology of constructive capriciousness. *Review of General Psychology, 5*, 270–290.

Silvia, P. J. (2005). What is interesting? Exploring the appraisal structure of interest. *Emotion, 5*, 89-102.

Sobel, D. (1996). *Beyond ecophobia: Reclaiming the heart in nature education.* Great Barrington, MA: The Orion Society and The Myrin Institute.

Stryker, S. (1980). *Symbolic interactionism: A social structural version*. Menlo Park, CA: Benjamin-Cummings.

Tan, E., & Barton, A. C. (2008). Unpacking science for all through the lens of identities-in-practice: The stories of Amelia and Ginny. *Cultural Studies of Science Education*, *3*(1), 43-71. Also available at http://barton.wiki.educ.msu.edu/file/view/tan+and+ACB.pdf

Tolman, E. C. (1932). *Purposive behavior in men and animals*. New York: Appleton-Century-Crofts.

Tucker-Raymond, E., Varelas, M., Pappas, C. C., Korzh, A., & Wentland, A. (2007). “They probably aren’t named Rachel”: Young children’s scientist identities as emergent multimodal narratives. *Cultural Studies of Science Education*, *1*(3), 559-592.

Tytler, R. (2007). Re-imagining science education: Engaging students in science for Australia. *Lab-talk*, *51*(3), 6-9. Also available at http://dro.deakin.edu.au/eserv/DU:30013528/tytler-reimaginingscienceeducation-2007.pdf

Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity*. Cambridge, UK: Cambridge University Press.

Wigfield, A., & Eccles, J. S. (1992). The development of achievement task values: A theoretical analysis. *Developmental Review*, *12*(3), 265-310.

Wigfield, A., Tonks, S., & Eccles, J. S. (2004). Expectancy value theory in cross-cultural perspective. In D. McInerney & S. van Etten (Eds.), *Research on sociocultural influences on motivation and learning* (pp. 165-198). Greenwich, CT: Information Age Publishers.

Wigfield, A., Tonks. S., & Klauda, S. L. (2009). Expectancy-value theory. In K. R.Wentzel & A. Wigfield (Eds.), *Handbook of motivation at school* (pp. 55-75). New York, NY: Routledge.

Chapter 2. Studies 1A, 1B & 1C: Designing and Validating a Survey for Measuring High School Students' Science Identities, Expectations of Success in Science, Values of Science and Environmental Attitudes (SIEVEA)

Study 1A – Designing SIEVEA Survey Instrument

Introduction. Students' science identities, expectations of success in science, and values of science play an important role in their science learning. Therefore, it is imperative to have instruments that can be used to measure these constructs in a simple, yet reliable way. This was the primary motivation behind this study, which aimed to design a survey that can be used to measure high school students' science identities, expectations of success in science and values of science. In addition to the aforementioned three constructs, the survey also measured students' environmental attitudes. Since environmental issues relate to both science learning and contemporary concerns, it felt natural to include questions related to environmental attitudes in the survey, especially given the ease of the survey administration. Additionally, the planned applications of the survey were related to the domains of biology and environmental science learning. Therefore, the ability to measure students' environmental attitudes was helpful while assessing their science identities and motivation in science. Lastly, given that students' interests in science subjects are related to their motivation in learning science, it made sense to capture students' science preferences in the survey.

The review of literature regarding current and past research uncovered that various instruments were developed for measuring either motivational constructs in science learning (Mubeen & Reid, 2014) or environmental literacy (Johnson & Manoli, 2011; Zecha, 2010). These findings are summarized in Table 2.1.

Table 2.1
Questionnaire or Interview based Studies of Motivational and Environmental Constructs

Study	Discussion
Blatt (2014)	Study measured 10 high school students' environmental identities while conducting qualitative analysis. ~ This study was appropriate for a small sample.
Glynn & Koballa (2006)	Developed Science Motivation Questionnaire (SMQ) designed for college and high school students. ~ The survey had only 20 questions; there were no questions about students' environmental attitudes.
Swarat, Ortony & Revelle (2012)	Administered questionnaire focused on students' interest in science in diverse instructional episodes (IE). ~ The questionnaire measured students' interest in specific biology topics.
Glynn, Brickman, Armstrong, & Taasoobshirazi (2011)	Updated Science Motivation Questionnaire (SMQ-II). ~ This questionnaire was the updated form of the SMQ. It added 5 more questions. Also, it had no questions pertaining environmental attitudes.
Johnson & Manoli (2011)	Developed 2-MEV questionnaire which measured adolescences' environmental attitudes. ~Survey did not have any constructs measuring students' motivation of science learning.
Eilam & Trop (2012)	Investigated the relationship between environmental attitudes and environmental behaviors of students and their parents. ~ Questionnaire was used in order to understand the processes that influence environmental behaviors and environmental attitudes. The questionnaire is focused on environmental education strictly and has no science education component measuring motivation in it.
Zecha (2010)	Analyzed students' environmental awareness. ~ Questionnaire was focused on assessing students' environmental knowledge, attitudes and actions.
Plantae, O'Keefe & Theoret (2012)	Explored the relationship between expectancy-value and achievement goal theories and their role

	in predicting achievement behaviors. ~ Two questionnaires assessed various motivational constructs including students’ expectancies and values in mathematics and language arts.
Bradley, Waliczek & Zajicek (1999)	Studied the relationship between environmental knowledge and environmental attitude of high school students. ~ Utilized questionnaires as a pre-test and post-test for assessing aforementioned constructs before and after intervention.
Fortus & Vedder-Weiss (2014)	Examined students’ continued motivation (CM) for science learning. ~ Survey was developed to measure CM constructs. The survey had only two items for assessing students’ environmental attitudes. It should be noted that the environmental attitudes construct was blended with science constructs and the second item had the same question as the first one; it was merely negatively worded.

As can be seen from the data in Table 2.1, existing instruments do not assess motivation in science learning and environmental attitudes in a single survey. Why is it useful to design a survey that can be used to measure these constructs together, rather using multiple surveys? In other words, is it really necessary to have yet another instrument? What useful data can it provide to science educators and science researchers?

First, the logistics of administering one combined survey is much easier than that of using multiple instruments. It is convenient for both the researcher and the study’s participants, and it involves reduced administrative overhead and paperwork. Secondly, a short survey reduces the possibility of participants getting bored and, as a result, skimming over questions or leaving the survey incomplete. Thus, it has a greater chance of being complete, accurate and candid than bigger surveys with lots of questions. Therefore, one can expect that collected data using this type of survey should be reasonably complete and accurate. Additionally, since there are no known instruments that measure science identities, motivational and environmental attitudes in one survey, the development of the current survey (SIEVEA) was necessary and practical. Next, this instrument is especially useful for researchers who study students' environmental attitudes in conjunction with their science learning motivation and science identities, as it allows for collection of pertinent data in a single data set using one combined instrument. Lastly, unlike Swarat, Ortony & Revelle's (2012) questionnaire, which measured students' interest in specific topics within a single domain (e.g., biology), and Matsui, Matsui, & Ohnishi's (1990) study, that asked students to rate how much they liked math, this instrument aimed at measuring students' interest in various science subjects.

Method. The survey designed in this study aimed at measuring high school students' science identities, expectations of success in science, values of science and environmental attitudes. The survey (see Appendix A) used a 5-point Likert scale developed by Likert (1932) for the measurement of attitudes. The survey was web-based and published on the http://www.qualtrics.com web site. In this survey, students were asked to indicate their agreement with each statement by selecting from the following choices: (1) Strongly Disagree, (2) Disagree, (3) Neither Agree nor Disagree, (4) Agree, and (5) Strongly Agree. The first two questions of the survey captured students' favorite subjects and gender. The remaining items were designed to measure students' science identities, their expectations of success in science classes, how they value science, and their environmental attitudes.

The developed survey items reflected on whether students viewed themselves as scientists, how students were viewed by others as scientists, students' expectations about being successful in science, how students valued science and whether students had positive attitudes and an appreciation of their environment. The number of items did not exceed fifteen, the survey questions were simplified and coherent, and the structure of the survey was set up with a multiple–choice format in order to resemble the most common test format familiar to students.

Survey construction. The current survey's design was based on existing literature on researched constructs and instruments for their measurement. The design incorporated modifications to existing instruments in order to accommodate the researcher's interests. The survey was reviewed by two researchers who made recommendations regarding the wording of some questions. The survey was revised accordingly. The development of items was guided by literature regarding science identities (Carlone & Johnson, 2007), expectation of success (Eccles et al., 1998), values of science (Eccles & Wigfield, 1995) and environmental attitudes (Dunlap & Van Liere, 1978; Dunlap, Van Liere, Mertig, & Jones, 2000).

Science identities were assessed by three items constructed according to the Carlone & Johnson (2007) 'Science Identity' initial model. This model includes three overlapping dimensions: competence, performance and recognition. For the current study, two dimensions were used: performance and recognition. The *expectation of success in science* construct was adopted from the Eccles et al. (1998) expectancy-value framework, as well as from Bandura's (1997) self-efficacy theory. Both theories accentuate the expectations of success as an integral part of motivation. The survey items assessing students' expectations of success in science were modeled after the works of Plante et al. (2013) and Glynn et al. (2011).

The *value of science* construct was adopted from the Eccles et al. (1998) expectancy-value framework. This construct is highly related to the importance of learning sciences (Eccles & Wigfield, 1995), as well as how individuals value an academic task (Plante et al., 2013). Questions used for measuring this construct were created based on MSLQ (Pintrich & DeGroot, 1990), SMQ (Glynn & Koballa, 2006), and the survey implemented by Eccles & Wigfield (1995). The value of science construct was assessed with four items, including two items measuring intrinsic value and two items measuring attainment value.

The *environmental attitudes* construct was developed after reviewing existing literature regarding measuring environmental attitudes (Dunlap & Van Liere, 1978; Dunlap, Van Liere, Mertig, & Jones, 2000). The Dunlap et al. (2000) revised New Ecological Paradigm (NEP) scale was used as a framework. Figure 2.1 depicts survey items and their categories.

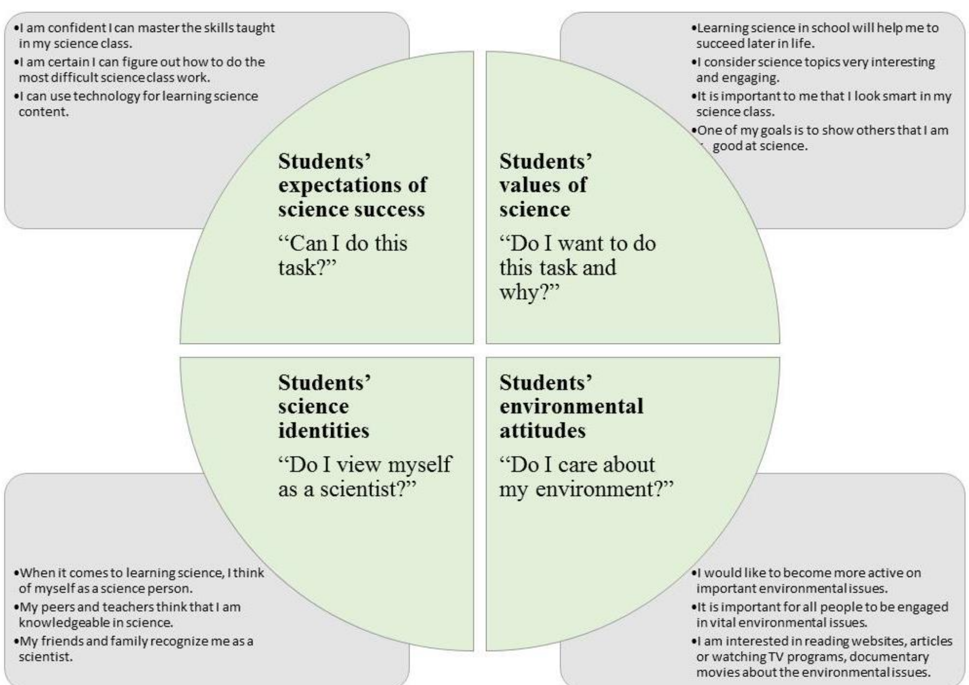


Figure 2.1. Survey Categories and Items

Since constructs like motivation of science learning and environmental attitudes are not directly observable, they are considered latent variables (Glynn et al., 2008) or composite variables (Thomson, 2004). Moreover, direct measurement of these latent variables is not possible. However, it is possible to indirectly measure these variables by using questionnaire items that act as empirical indicators of how the constructs are conceptualized by students (Glynn et al., 2008). This process of representing not directly observable constructs as composite variables made of easily measurable questionnaire items is known as operationalization (Agarwal, 2011). Table 2.2 summarizes how this study's constructs were mapped to questionnaire items and what literature and existing instruments were used in support of the mappings.

Table 2.2
Instrument Items and Developed Constructs Matrix

Item	Construct	How the item was developed
3. Learning science in school will help me to succeed later in life.	Values of Science	This question assesses the intrinsic value of science (Eccles & Wigfield, 2002). It is modeled after items from MSLQ (Pintrich & DeGroot, 1990).
4. I am confident I can master the skills taught in my science class.	Expectations of Success in Science	Modeled after items from Plante et al. (2013) questionnaire and Glynn et al. (2011) SMQ-II.
5. I consider science topics very interesting and engaging.	Values of Science	This question assesses the intrinsic value of science (Eccles & Wigfield, 2002). It is modeled after items from MSLQ (Pintrich & DeGroot, 1990) and Eccles & Wigfield (1995).
6. When it comes to learning science, I think of myself as a science person.	Science Identity	Science Identity Model designed by Carlone & Johnson (2007). The question assesses the model's

		‘recognition’ dimension.
7. My peers and teachers think that I am knowledgeable in science.	Science Identity	Science Identity Model designed by Carlone & Johnson (2007). The question assesses the model’s ‘recognition’ and ‘performance’ dimensions.
8. I am certain I can figure out how to do the most difficult science class work.	Expectations of Success in Science	Modeled after items from Plante et al. (2013) questionnaire and Glynn et al. (2011) SMQ-II.
9. I can use technology for learning science content.	Expectations of Success in Science	Modeled after items from Glynn et al. (2011) Science Motivation Questionnaire-II (SMQ-II).
10. My friends and family recognize me as a scientist.	Science Identity	Science Identity Model designed by Carlone & Johnson (2007). The question assesses the model’s ‘recognition’ dimension.
11. It is important to me that I look smart in my science class.	Values of Science	This question assesses the attainment value of science (Eccles & Wigfield , 1995). It was modeled after items from SMQ (Glynn & Koballa, 2006) questionnaire.
12. I would like to become more active on important environmental issues.	Environmental Attitudes	Modeled after survey items from Stern, Powel, & Ardoin (2011). This survey combined all questions assessing students’ environmental concerns.
13. One of my goals is to show others that I am good at science.	Values of Science	This question assesses the attainment value of science. It is researcher-developed and is based on the existing literature and expectancy-value theory.
14. It is important for all people to be engaged in vital environmental issues.	Environmental Attitudes	Modeled after survey items from DiEnno & Hilton (2005).
15. I am interested in reading website, articles or watching TV programs, documentary movies about the	Environmental Attitudes	This item is researcher-developed. It is based on the revised New Ecological Paradigm (NEP) scale.

environmental issues.

The Lexile Framework for Reading system (Meta-Metrics, Inc., 2012) was used to analyze the readability of survey's questions. The analyses showed that the survey's reading complexity did not exceed the reading ability of a typical high-school (grade 9) student. The use of age appropriate questions also enhanced the instrument's validity (Meta-Metrics, Inc., 2012). Moreover, scholars stress the importance of simplicity in wording and avoidance of ambiguous language when designing surveys (Gehlbach & Brinkworth, 2011).

Instrument piloting. Before conducting pilot study, survey's face validity was measured. Although face validity is considered a weak instrument for gauging survey's validity, it can tell whether the purpose of the instrument is clear or not (Nevo, 1985). The survey was shown to the layperson in order to verify the instrument's face validity (Rea & Parker, 2014).

The survey was piloted with 30 non-science major students. Pilot testing data was analyzed in order to identify and remedy problematic items. First, Pearson correlation coefficients were calculated for all pairs of items (13 items, 78 pairs in all) in order to assess inter-item correlations. Additionally, the frequency charts of student answers were built.

Careful examination of Pearson coefficients showed that all items except one were correlated to one or more item(s) in their construct group (e.g., items #14 and #15, which were both environmental attitude items, had a Pearson coefficient of 0.476). However, item #12 ("Learning about environment in school is important for me") did not correlate strongly with any other item (all coefficients were 0.255 or less). Therefore, this item was replaced with a different environmental attitude item. Similarly, analysis of answer frequencies showed that no student disagreed with item #10 (“Hands-on activities and lab make studying science fun”). This item was also replaced. Figure 2.2 depicts the main steps of SIEVEA’s development.

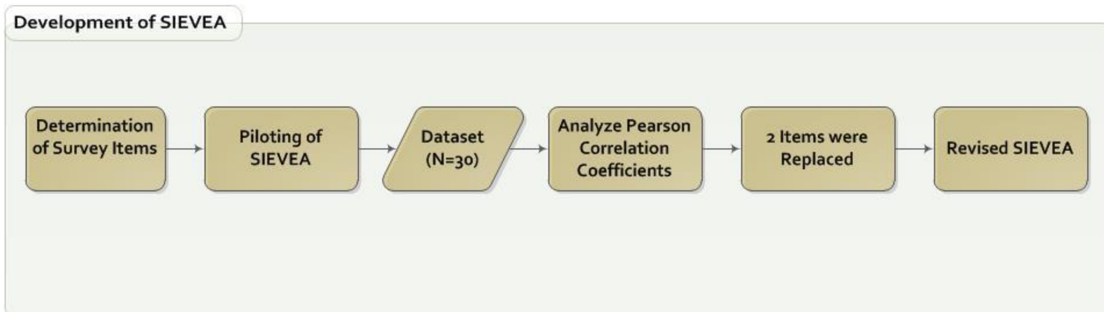


Figure 2.2. Development of SIEVEA

Prior to conducting this study with a full sample of students, a professor who had expertise in theories of motivation and science identities was consulted for reviewing the motivational and identity constructs and critiquing the wording of survey items. Two items were modified based on professor’s feedback.

Data collection. The data collection took place during the 2014-2015 (May-June) and 2015-2016 (September-January) academic years. The survey was very easy to administer; it was online and took less than 5 minutes to complete on average (M = 3.4 minutes after removing some abnormally high response times that were greater than 15 minutes, as some students took a break while doing the survey). Students needed to agree to the electronic consent form, which was approved by the Institutional Review Board (IRB), in order to proceed with completion of the survey. The collected data was downloaded from the Qualtrix website and was saved on a password-protected computer for data analysis.

Participants. A total of 1,911 high school students from grades nine to twelve (14 to 17 years old) took the SIEVEA. However, only 1,764 students' responses were complete and used for data analysis. Of these participants, 930 were girls and 827 were boys. Seven students chose not to report their gender. Participating students were from 11 school districts in New Jersey, Pennsylvania and Connecticut. Each school district received a unique hyperlink, which was distributed among its high school students. The entire student population was asked to participate. Participation was completely voluntary. The participant schools were urban, suburban and private high schools. Other than gender, no other demographic data was collected. The survey's data indicated a 92.3% completion rate.

Exploratory factor analysis. The only directly observable measures in the instrument were the students' gender and favorite science subject. The remaining constructs assessing students' attitudes and motivation in science were latent variables that could not be observed or measured directly. Not all responses were included in data analysis. Responses with one or more missing answers were excluded. Because of this criterion of exclusion, the final data set contained 1,660 responses out of the initial 1,764 valid responses. Thus, 104 responses (5.9% of total) were excluded.

Prior to doing factor analysis, the data in this survey was analyzed using descriptive statistics. First, item level statistics were collected for all survey items related to the four variables of interest (see Table 2.3).

Table 2.3
Summary of Item Level Descriptive Statistics

Item	Range	Mean	SD	Skewness	SE	Kurtosis	SE
3	1 to 5	3.88	.980	-.718	.058	.112	.117

4	1 to 5	3.89	.847	-.792	.058	.867	.117
5	1 to 5	3.79	1.013	-.804	.058	.272	.117
6	1 to 5	3.18	1.179	-.178	.058	-.833	.117
7	1 to 5	3.6	.928	-.445	.058	-.049	.117
8	1 to 5	3.33	1.050	-.434	.058	-.387	.117
9	1 to 5	4.08	.798	-.985	.058	1.611	.117
10	1 to 5	2.52	1.110	.405	.058	-.502	.117
11	1 to 5	3.27	1.075	-.268	.058	-.469	.117
12	1 to 5	3.41	.998	-.339	.058	-.228	.117
13	1 to 5	3.08	1.103	-.133	.058	-.688	.117
14	1 to 5	3.65	.949	-.515	.058	.056	.117
15	1 to 5	3.33	1.083	-.409	.059	-.440	.117

According to data in Table 2.3, responses ranged from 1 to 5. Mean scores of all items ranged from 2.52 to 4.08. Out of these items, 12 items were negatively skewed (ranging between -.985 and -.133). Item #10 was positively skewed (with skewness of .405). Kurtosis statistics ranged from -.833 to 1.611 with a standard error of .117. Item #9 ("I can use technology for learning science content.") had the largest skewness (-.985) and the largest kurtosis (1.611) by absolute value. This showed that Item #9's data had a significant non-normality.

Next, the 13x13 matrix of Pearson correlation coefficients between all survey items was computed (see Table 2.4). All Pearson correlation coefficients were significant at $p < .01$ (2-tailed).

Table 2.4
Intercorrelations for SIEVEA Items

Item	3	4	5	6	7	8	9	10	11	12	13	14	15
3. Learning science in school will help me to succeed later	–												

in life.													
4. I am confident I can master the skills taught in my science class.	.47	–											
5. I consider science topics very interesting and engaging.	.54	.45	–										
6. When it comes to learning science, I think of myself as a science person.	.56	.48	.67	–									
7. My peers and teachers think that I am knowledgeable in science.	.38	.51	.39	.48	–								
8. I am certain I can figure out how to do the most difficult science class work.	.37	.52	.40	.49	.50	–							
9. I can use technology for learning science content.	.26	.30	.26	.26	.26	.29	–						
10. My friends and family recognize me as a scientist.	.46	.38	.50	.65	.40	.45	.18	–					
11. It is important to me that I look smart in my science class.	.30	.25	.27	.32	.35	.29	.21	.33	–				
12. I would like to become more active on important environmental issues.	.29	.18	.33	.33	.16	.18	.20	.28	.27	–			
13. One of my goals is to show others that I am good at science.	.45	.29	.41	.49	.29	.32	.16	.47	.52	.40	–		
14. It is important for all people to be engaged in vital environmental issues.	.29	.14	.31	.28	.18	.15	.21	.25	.25	.56	.33	–	
15. I am interested in reading websites, articles or watching TV programs, documentary movies	.28	.19	.39	.38	.18	.23	.21	.34	.23	.50	.34	.45	–

about the environmental issues.

Note. All coefficients are significant at $p < .01$ (2-tailed).

As data in Tables 2.4 showed, even though Item #9 (“I can use technology for learning science content.”) had statistically significant correlations with other items in its group ($p < .01$), its Pearson correlation coefficients were noticeably smaller than those of others (the greatest coefficient was .30). Therefore, both item’s descriptive statistics from Table 2.3 and its correlation coefficients from Table 2.4 indicated that this item was somewhat out of place and its usefulness and place within the instrument was questionable. Moreover, a quick test-run of the factor extraction process on survey’s data showed that Item #9 did not factor well into any extracted factor: its factor loadings were low (about .300).

These results indicated that this item (“I can use technology for learning science content.”) did not fit well within other items of the survey and the four researched constructs. Therefore, it made sense to drop this item from the instrument and exclude its data from subsequent analyses. After Item #9 was removed, an exploratory factor analysis (Gorsuch, 2003) was conducted on the remaining 12 items. Exploratory factor analysis (EFA) was a suitable method for the current study, since it is used when the links between the observable and unobservable (latent) variables are uncertain (Byrne, 2013).

Detailed analysis of past studies involving EFA indicated that many researchers failed to report important information concerning factor extraction procedures, sample size, number of measured variables, determination of the number of factors for extraction, the type of factor rotation and factor scores (Russell, 2002). Because of this, it was the researcher’s priority to ensure that all necessary parameters and procedures for conducted EFA were addressed in this study. Hence, exploration of the data's factor structure was completed based on research recommended guidelines (Fabrigar, Wegener, MacCallum and Strahan, 1999; Matsunaga, 2010; Russell, 2002).

It should be noted that EFA has lots of similarities with Principal Component Analysis (PCA) since both belong to the family of dimension reduction statistical procedures. This has caused confusion and led some researchers to misuse PCA in their studies instead of implementing EFA (Henson & Roberts, 2006). PCA is used to reduce the number of observed variables to a smaller number of principal components that account for most of the variance of the observed variables, whereas EFA is used to identify the number of latent constructs and the underlying factor structure of observed variables. Moreover, PCA assumes that the observed items were measured without measurement errors, whereas EFA does not make such an presumption regarding the data (Matsunaga, 2010).

EFA was appropriate for this study, because EFA's main purpose is "… to generate hypotheses by identifying, describing and classifying data" (Child, 2006, p.108). Likewise, EFA helps the researcher to "… identify a set of unobserved (latent) factors that reconstruct the complexity of the observed (manifest) data" (Matsunaga, 2010, p. 98). The data sample of the current study was an "excellent" sample, as it had more than 1,000 subjects and was therefore appropriate for EFA (Comrey & Lee, 1992). EFA was carried out using the SPSS (Version 24) statistical package. Factor analysis allows researchers to reveal relationships between the items by examining their factor loadings, where factors are defined as constructs or dimensions (Kline, 2014). Even though the survey developed in this study is not large, factor analysis was successfully conducted in past studies with as few as 15 item instruments (Matsui et al., 1990).

Prior to conducting factor analysis, item-to-item correlation was examined by conducting the Kaiser-Mayer-Olkin (KMO) test and Bartlett's test for Sphericity. The statistical value of KMO ranges from 0 to 1: values close to 0 indicate the data is not suitable for factor analysis, whereas values close to 1 demonstrate that factor analysis should produce well-defined and valid factors. Moreover, Kesier (1974) considered values greater than 0.5 appropriate for applying factor analysis. Additionally, Bartlett's test determines whether the R-matrix is an identity matrix or not. In other words, this test shows whether any correlations between variables exist (Field, 2005). Bartlett's test should be significant ($p < .05$) in order for factor analysis to be appropriate.

Both the KMO test and Bartlett's test for Sphericity indicated that the correlations matrix was appropriate for conducting factor analysis (Glyn et al., 2008). Indeed, the Kaiser-Mayer-Olkin Measure of Sampling Adequacy of data was .893 (close to 1). Likewise, Bartlett's test of Sphericity produced $p < .001$. The choice of the most suitable method for exploratory factor analysis was based on the recommendation of Fabrigar, Wegener, MacCallum and Strahan (1999). Fabrigar et al. (1999) argued that the maximum likelihood method should be used for relatively normally distributed data, whereas the principal factor methods are a better choice for data with significant departures from normality. The descriptive statistics (see Table 2.3) indicated that item level data's distributions did not significantly depart from normal distribution. For this reason, the maximum likelihood method was used for factor extraction.

Also, Varimax orthogonal rotation was chosen, as Varimax tends to maximize variance by making high factor loadings higher and low factor loadings lower (Tabachnik & Fidell, 2001). Moreover, Varimax rotation produces a simple structure which makes the process of interpretation much easier (Glynn et al., 2008).

Factor extraction. When implementing EFA, one of the most important decisions to make is the number of factors to retain. Researchers use several methods for making this decision. Kaiser's (1960) eigenvalue-greater-than-one rule is the most commonly used method for deciding which factors should be retained (Fabrigar et al., 1999). This rule simply states that factors with eigenvalues greater than one should be retained, while those with eigenvalues less than or equal to one should be discarded.

For the first attempt, the principal axis factoring method was used and factors were extracted using Kaiser's eigenvalue-greater-than-one rule. This resulted in 2 factors with eigenvalues 5.116 and 1.560. Careful examination of the scree plot showed that the next eigenvalue (.931) was close to Kaiser's cutoff value of 1, so it may make sense to retain more than 2 factors (see Figure 2.3).

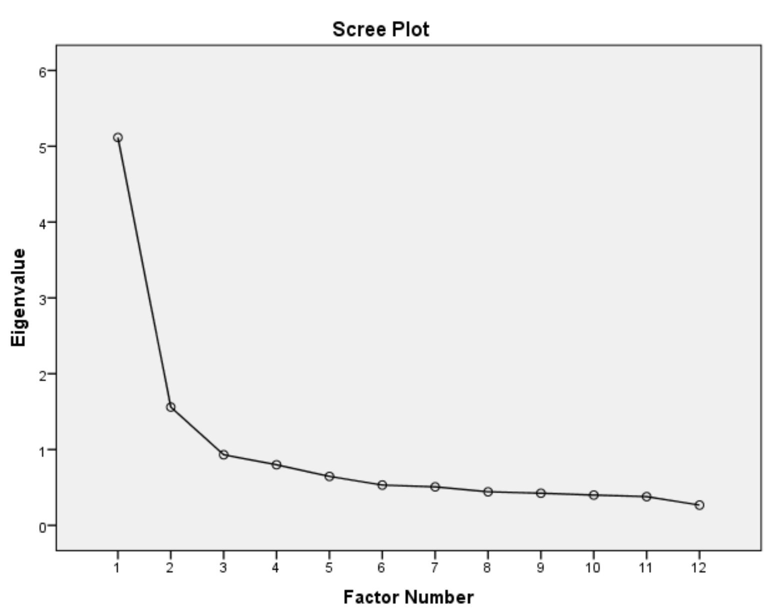


Figure 2.3. Scree Plot of Principal Axis Factoring

While Kaiser's method is very simple and easy to use, it has some drawbacks (Fabrigar et al., 1999). To begin with, Kaiser's method was originally proposed for the principal component analysis, which uses a correlation matrix that is different from the matrix used for EFA: unities at the diagonal instead of communality estimates. Another problem is this method's cutoff value of 1. Indeed, why would one retain a factor with an eigenvalue of 1.001, but ignore one with an eigenvalue of .999? The difference is so insignificant that any decision based solely on this seems arbitrary. Additionally, various studies showed that Kaiser's rule tends to significantly overestimate or underestimate the number of factors (Zwick & Velicer, 1986).

Since it was not obvious how many factors the model should contain, it was necessary to implement additional analysis, as suggested by numerous methodologists (Costello & Osborne, 2011).

Parallel analysis. Horn (1965) proposed an alternative method for factor retention. His approach, Parallel Analysis (PA), is essentially a Monte Carlo simulation process. It computes a new set of eigenvalues of the correlation matrix using randomly generated dataset with the same numbers of observations and variables as the original data. Then it compares these eigenvalues to the ones computed by EFA. If the eigenvalues based on the random data are greater than the matching eigenvalues from EFA, then one can safely assume that corresponding EFA factors are not significant and can be dropped from the model. Otherwise, the factors are retained.

Many studies confirmed that PA is the best method for determining the number of factors to be retained during EFA (Zwick & Velicer, 1986). Its results are more accurate and reliable than those produced by other methods, including Kaiser’s method (Glorfeld, 1995). Although PA is a reliable and accurate method for factor selection, this method was not widely used in research (Fabrigar et al., 1999). Instead, other less reliable methods like the Kaiser’s eigenvalue-greater-than-one rule were extensively used. One reason for this is that popular statistical packages, like SPSS, provide Kaiser's rule as the default option for factor retention, whereas PA is not built in into these packages and requires special programming using statistical languages. Fortunately, one does not have to write his/her own code to do PA since some freely available SPSS macros and other free statistical packages are available for doing this.

For the current research SPSS program rawpar.sps developed by O'Connor (2000) was used to conduct PA. This program can run PA on either normally distributed randomly generated data or on permutations of the original raw data set. Using the original raw data allows preserving the distributions of the original raw variables in the permuted versions of data thus making the results of PA highly accurate. For that reason, the original raw data set option was used (compute randtype = 2; for permutations of the raw data set). Next, compute kind parameter was set to 2 for principal axis/common factor analysis. Lastly, ndatsets parameter was changed from default 100 to 1,000 in order to do 1,000 permutations which is considered sufficient number of iterations to produce reliable results (Matsunaga, 2010).

PA calculations produce 3 sets of data: raw data eigenvalues, means of eigenvalues of random data and upper limits of 95% confidence interval for means (see Table 2.5). Additionally, PA procedure produces a scree plot of these 3 data sets (see

Figure 2.4). This scree plot is easy to inspect in order to identify the number of factors for extraction.

Table 2.5
Raw Data Eigenvalues, Means and 95% Confidence Intervals

Root	Raw Data	Means	95% CI
1	4.580302	.148047	.181574
2	.965497	.110805	.134205
3	.336275	.083029	.103615
4	.248886	.058374	.077245
5	.091173	.035754	.053504
6	-.038037	.014634	.031484
7	-.063864	-.006338	.009386
8	-.120007	-.027078	-.012117
9	-.146232	-.048498	-.032766
10	-.154811	-.070770	-.053421
11	-.191673	-.094532	-.075383
12	-.202074	-.123776	-.100304

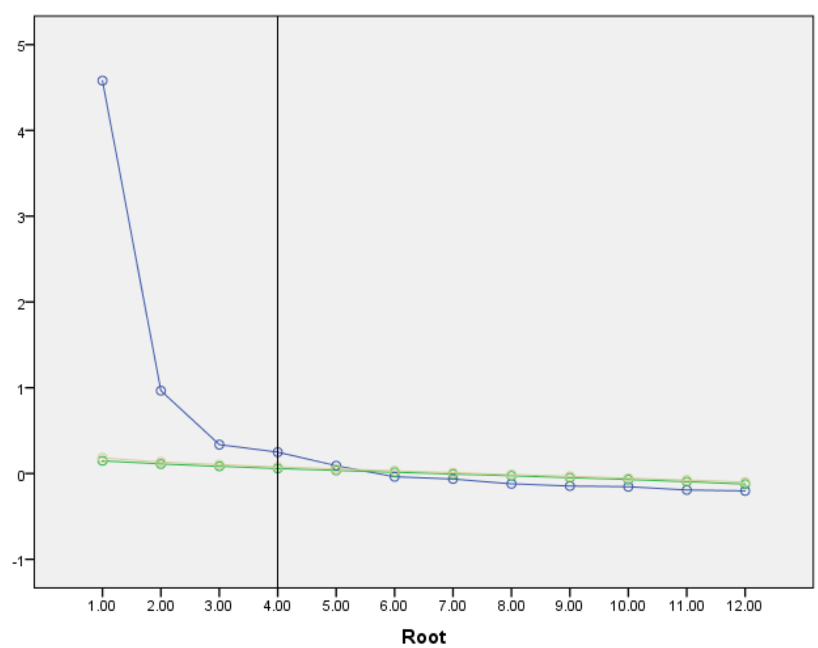


Figure 2.4. Scree Plot of Parallel Analysis

The results of PA showed that up to 4 factors can be extracted. Indeed, first 4 eigenvalues computed using the sample data are considerably greater than the upper limits of 95% confidence intervals of eigenvalues computed using randomly generated data (see Table 2.5 and Figure 2.4). Therefore, two more extractions were done: for the second attempt, the maximum likelihood method was used while forcing the extraction of 3 factors and, for the third attempt, 4 factors were extracted.

Candidate models. The factor extraction resulted in three candidate models: 2-factor model, 3-factor model and 4-factor model. Table 2.6 summarizes variance explained before and after Varimax rotation for each model according to research recommendation (Russell, 2002).

Table 2.6
Variance Explained by the Unrotated and Varimax Rotated Factors

	2-Factor Model		3-Factor Model		4-Factor Model	
Factor	Unrotated	Rotated	Unrotated	Rotated	Unrotated	Rotated
1	38.30	29.45	38.68	27.56	38.67	16.09
2	8.78	17.63	8.69	16.18	8.07	15.45
3	n/a	n/a	4.66	8.30	5.64	14.37
4	n/a	n/a	n/a	n/a	3.51	9.97
Total Variance	47.08	47.08	52.04	52.04	55.89	55.89

Table 2.7 contains the factor loadings (after Varimax rotation was applied) for all three candidate models. Most survey items had reasonably good loadings in all three candidate models. However, a few items' loadings on some factors were close indicating certain ambiguity in choosing of the factor they should belong to.

Table 2.7
Factor Loadings of SIEVEA Models Using Maximum Likelihood and Varimax Rotation

	2-Factor Model	3-Factor Model	4-Factor Model

Item	C1	C2	C1	C2	C3	C1	C2	C3	C4
3	**.618**	.285	**.605**	.268	.152	.369	**.449**	.216	.261
4	**.653**	.076	**.630**	.054	.146	**.670**	.257	.087	.109
5	**.658**	.341	**.686**	.352	.023	.347	**.607**	.289	.124
6	**.775**	.309	**.794**	.304	.087	.380	**.768**	.212	.183
7	**.626**	.076	**.582**	.028	.277	**.667**	.216	.075	.148
8	**.640**	.095	**.607**	.063	.201	**.620**	.267	.085	.166
10	**.663**	.279	**.650**	.260	.164	.345	**.550**	.188	.261
11	**.373**	.304	.246	.197	**.751**	.300	.090	.209	**.494**
12	.127	**.774**	.120	**.753**	.148	.068	.135	**.748**	.188
13	**.475**	.431	.409	.373	**.437**	.140	.320	.252	**.801**
14	.114	**.693**	.104	**.677**	.140	.088	.098	**.693**	.140
15	.238	**.605**	.247	**.613**	.054	.108	.267	**.588**	.102

The following discussion summarizes each model along with its factor structure in order to balance the number of retained factors with interpretability of the model.

Two-factor model. The two-factor model accounted for 47.08% of the variability in the intercorrelation matrix. Item loadings are shown in Table 2.7.

The first factor included 9 items (Items #3-8, 10, 11, 13). The loadings of these items on the first factor ranged between .373 and .775. All these items reflected students' science learning related motivational and self-efficacy constructs: how students value science (e.g., "It is important to me that I look smart in my science class."), if students expect to succeed in science learning (e.g., "Learning science in school will help me to succeed later in life."), if students view themselves as a "science person" (e.g., "My friends and family recognize me as a scientist."). The second factor included 3 items (Items #12, 14, 15). The loadings of these items on the second factor ranged between

.605 and .774. All these items reflected students' environmental attitudes (e.g., "It is important for all people to be engaged in vital environmental issues").

This model clearly separated science learning related constructs and environmental attitudes into two different factors. The first factor (C1) was named "Science Learning Motivation and Self-Efficacy." The second factor (C2) was named "Environmental Attitudes."

Three-factor model. The three-factor model accounted for 52.04% of variability within the intercorrelation matrix. Item loadings are shown in Table 2.7.

The first factor included 7 items (items #3-8, 10). The loadings of these items on the first factor ranged between .582 and .794. All these items reflected students' science identities (e.g., "When it comes to learning science, I think of myself as a science person") and their expectations of success in science learning (e.g., "I am certain I can figure out how to do the most difficult science class work").

The second factor included 3 items (items #12, 14, 15). These items' loadings on the third factor were .753, .677 and .613, respectively. All these items reflected students' environmental attitudes (e.g., "I would like to become more active on important environmental issues.").

The third factor included 2 items (items #11, 13). These items' loadings on the thrid factor were .751 and .437 respectively. These items reflected students' science values ("It is important to me that I look smart in my science class" and "One of my goals is to show others that I am good at science").

This model retained the environmental attitude factor (C2) while decomposing the science learning related factor into two different factors (C1, C3). The first factor was named “Science Identities and Motivation.” The second factor was named “Environmental Attitudes.” The third factor was named “Science Values.”

Four-factor model. The four-factor model accounted for 55.89% of variability within the intercorrelation matrix. Item loadings are shown in Table 2.7.

The first factor included 3 items (items #4, 7, 8). The loadings of these items on the second factor ranged between .620 and .670. These items reflected students’ expectations of success in science learning (e.g., “I am confident I can master the skills taught in my science class”). The second factor included 4 items (items #3, 5, 6, 10). The loadings of these items on the second factor ranged between .449 and .768. All these items reflected students’ science identities (e.g., “When it comes to learning science, I think of myself as a science person”). The third factor included 3 items (items #12, 14, 15). These items’ loadings on the third factor were .748, .693 and .588, respectively. All these items reflected students’ environmental attitudes (e.g., “I would like to become more active on important environmental issues.”). The fourth factor included 2 items (items #11, 13). These items’ loadings on the fourth factor were .494 and .801 respectively. These items reflected students’ science values (“It is important to me that I look smart in my science class” and “One of my goals is to show others that I am good at science”).

This model retained the environmental attitude factor (C3) while decomposing the science learning related factor into three different factors (C1, C2, C4), one per each construct. The first factor was named “Expectations of Success in Science Learning.” The second factor was named “Science Identities.” The third factor was named “Environmental Attitudes.” The fourth factor was named “Science Values.”

Table 2.8 summarizes factors' eigenvalues, percentage of variance and Cronbach's alpha for all three models. The Cronbach's alpha is the most common metric for measuring an instrument's reliability. Alpha values between .70 and .95 are considered "good" (DeVellis, 2003) with an alpha greater than .80 considered "very good" according to the guidelines provided by DeVellis (2012). For the current study, the Cronbach's alpha was measured for each factor and it ranged from .684 to .870 (see Table 2.8). Cronbach's alpha values for factors in two-factor model were .870 and .749, whereas for three-model and four-factor model they ranged from .684 to .864 and .684 to .837 respectively. These numbers indicated a good level of the survey's internal consistency, namely, a close relationship among survey's items grouped by factors. This measurement provided an additional proof of survey's reliability.

The proposed study is repeatable and its results can be replicated. Therefore, it aligns well with Joppe's (2000, p. 1) definition of quantitative research's reliability, which is stated as "consistent over time" and "the results can be reproduced under a similar methodology." Likewise, carrying out the study in diverse settings (9th – 12th grade levels in urban, suburban, private schools in New Jersey, Pennsylvania and Connecticut) and receiving consistent results enhanced the external validity.

Table 2.8

Factors' Eigenvalue, Percent of Variance Explained, and Cronbach's Alpha

Factor	Eigenvalue	% of variance	Cumulative %	Cronbach's alpha
2 Factor Model				
C1	5.116	29.45	29.45	.870
C2	1.560	17.63	47.08	.749
3 Factor Model				

C1	5.116	27.56	27.56	.864
C2	1.560	16.18	43.74	.749
C3	.931	8.30	52.04	.684
4 Factor Model				
C1	5.116	16.09	16.09	.753
C2	1.560	15.45	31.55	.837
C3	.931	14.37	45.92	.749
C4	.799	9.97	55.89	.684

Figures 2.5, 2.6 and 2.7 visually display the percent of variance in the data contributed to each factor for two-factor, three-factor and four-factor models.

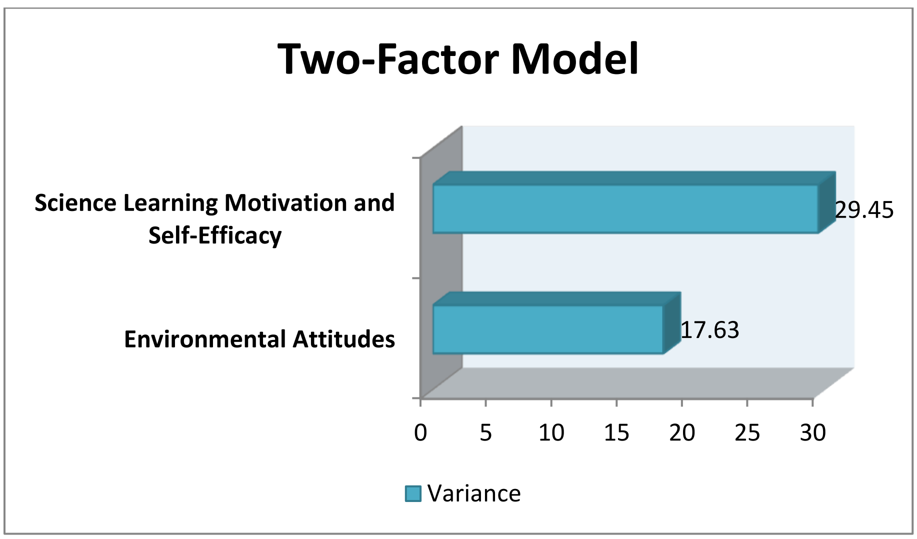


Figure 2.5. Percent of Variance by Each Factor: Two-Factor Model

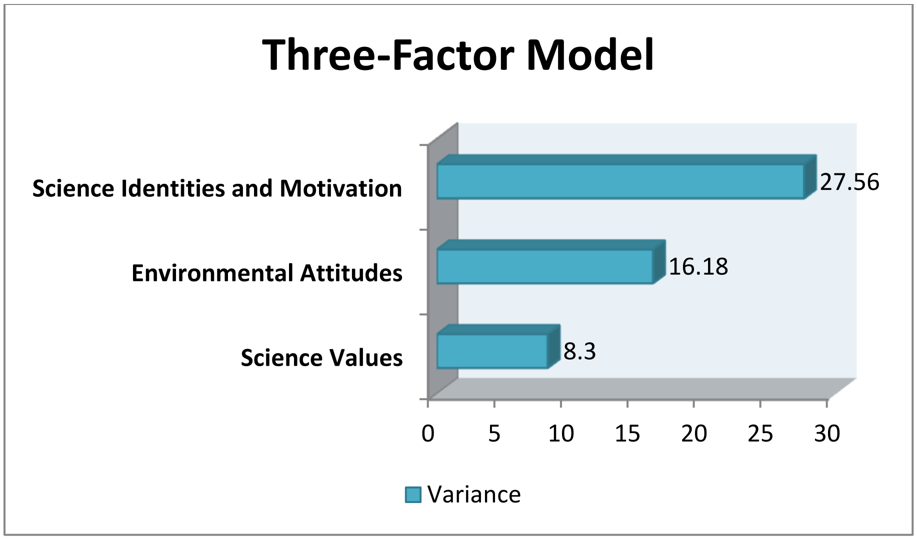


Figure 2.6. Percent of Variance by Each Factor: Three-Factor Model

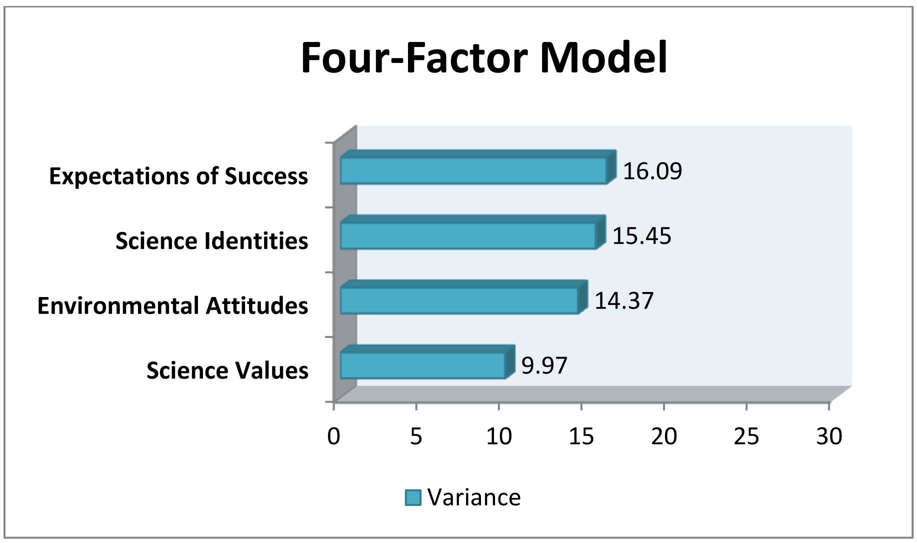


Figure 2.7. Percent of Variance by Each Factor: Four-Factor Model

The following analyses and dicussion cover model-fit evaluation and the final model selection. But prior to undertaking these analyses, it is necessary to discuss how model-fit evaluation is done and what are the related fit indexes.

Model-fit evaluation. In order to evaluate the model's fit to data, it is necessary to compute appropriate fit statistics and compare them to established cutoff values (Gignac, 2009). The chi-square statistic of the implied model is the simplest fit statistic. This statistic, also known as the "exact fit index," is easy to calculate and interpret. It should be noted that, unlike many statistical tests, the chi-square statistic test of model fit does not aim to reject the null hypothesis, meaning it is better not to have a statistically significant result (Thompson, 2004).

Unfortunately, the chi-square statistic as a fit index is not very reliable; it inclines to produce significant results as the sample size grows. Therefore, models with large sample sizes tend to have large and significant chi-squares (Thomson, 2004). This means that, when the chi-square statistic is computed, the sample size should be taken under consideration. Because of this statistic's high sensitivity to the sample size, researchers have largely stopped relying on it for evaluating the model's fit. Instead, a host of close-fit indexes was developed (Gignac, 2009).

Commonly used close-fit indexes are absolute close-fit indexes and incremental close-fit indexes (Hu & Bentler, 1999). Absolute close-fit indexes only implement statistics of the implied model, whereas incremental close-fit indexes are based on information from both the implied model and the null model (which assumes absence of any factors within the correlation matrix). For absolute close-fit, this study used *the Root Mean Square Error of Approximation* (RMSEA; Browne & Cudeck, 1993) and *the Standardized Root Mean Residual* (SRMR; Bentler, 1995). The main incremental close-fit indexes used in this study were *the Normal Fit Index* (NFI; Bentler & Bonnett, 1980), *the Tucker-Lewis Index* (TLI; Tucker & Lewis, 1973), and *the Comparative Fit Index* (CFI; Bentler, 1990).

The root mean square error of approximation (RMSEA) sidesteps the issue of the sample size by examining the closeness of the implied model to the data, rather than comparing it to the null model (Matsunaga, 2010; Steiger & Lind, 1980). Consequently, RMSEA's calculation uses only the implied model's parameters: the chi-squared statistic, the degrees of freedom and the sample size (Gignac, 2009).

$$RMSEA = \sqrt{\frac{\chi^2_{Implied} - df_{Implied}}{(N-1) * df_{Implied}}}$$

Different RMSEA cutoff values for model fit were proposed. Browne (1990) argued for RMSEA values less than or equal to .05 to indicate a close fit and anything below .08 for a reasonable fit. Reise, Widaman, & Pugh (1993) agreed with Browne (1990) in suggesting that RMSEA's acceptable values should not exceed .08, with values .05 or less specifying a close fit of a model. Marsh et al. (1988) also suggested .08 as an upper threshold for an acceptable fit. Thomson (2004) and Hu & Bentler (1999) suggested .06 or less as a good fit. Overall, the RMSEA values less than .05 indicate a close model fit, values between .05 and .08 are adequate for a good fit and anything greater than .10 is a sign of a poor fit (Browne & Cudeck, 1993).

The standardized root mean residual (SRMR) is the square root of the mean of the squared residuals of the residual correlation matrix produced by factor analysis (Bentler, 1995). This index is the standardized version of the root mean square residual (RMR) index. Though both indexes can be used for measuring the model's fit, SRMR is the most widely implemented residual-based index since it is easier to interpret (Matsunaga, 2010). The direct calculations of SRMR are somewhat complicated. Fortunately, the statistical packages and tools like SPSS provide facilities for calculating SRMR. The SRMR values of .08 or less suggest an acceptable model fit (Gignac, 2009; Hu & Bentler, 1999).

The normal fit index (NFI) looks at the difference between the chi-squared values of the implied model and the null model (Bentler & Bonnett, 1980). This index simply computes the ratio of the chi-squared values (see the formula below; Gignac, 2009) and is sensitive to sample size.

$$NFI = \frac{(\chi^2_{Null} - \chi^2_{Implied})}{\chi^2_{Null}}$$

NFI values greater than .95 are regarded as preferable for the model's fit (Thompson, 2004). According to Reise, Widaman, & Pugh (1993), .90 is a practical lower limit for indication of a satisfactory model fit.

The Tucker-Lewis index (TLI), also known as the non-normed fit index (NNFI), compares the fit of the implied model to that of the null model, assuming no relations between variables (Eccles & Wigfield, 1995). TLI can be computed using the following formula (Gignac, 2009):

$$TLI = \frac{(\chi^2_{Null}/df_{Null}) - \left(\chi^2_{Implied}/df_{Implied}\right)}{\left[(\chi^2_{Null}/df_{Null}) - 1\right]}$$

The conventional model fit cutoff value for TLI is .90 (Russell, 2002). However, some researchers suggested a more stringent fit-index value of .95 (Hu & Bentler, 1999).

The comparative fit index (CFI) assesses the discrepancy between the data and the implied model using adjusted chi-squared values in order to avoid sample size related issues prevalent in the chi-squared test of model fit and in NFI (Thompson, 2004). The following formula for calculating CFI was borrowed from Gignac (2009).

$$CFI = 1 - \frac{\left(\chi^2_{Implied} - df_{Implied}\right)}{\left(\chi^2_{Null} - df_{Null}\right)}$$

Similar to the TLI cutoff value, .90 or higher is considered a reasonable fit value for CFI (Bentler & Bonett, 1980; Russell, 2002). Hu & Bentler (1999) proposed to increase this cutoff value to .95, while other researchers suggested a less demanding value of .80 for an acceptable fit (Bollen, 1989).

There are other absolute and incremental close-fit indexes such as the goodness of fit index (GFI), the adjusted goodness of fit index (AGFI), and Bollen's incremental fit index (IFI) (Baumgartner & Hombur, 1996; Bollen, 1989). However, these indexes are not wildly used; therefore, this study and subsequent PCFA and CFA analyses considered only the following common fit indexes when accessing models' fit: RMSEA, NFI, TLI, CFI and SRMR. These indexes assess different aspects of a model fit (Kline, 2005). Therefore, using all of them allows to make accurate and reliable conclusions about the model's fit.

Model selection. The model selection in the final step of EFA. In order to pick the final model out of three candidate models, the following three factors were considered: 1) models' fit to data, 2) how well did each model align with the research constructs, and 3) models' factor loadings.

First, all three candidate models were evaluated in regard to their fit to data. During the factor extraction, using the maximum likelihood method, Bartlett's test of sphericity option was chosen. This produced the chi-square statistics and the degrees of freedom of the null model. Additionally, the maximum likelihood method calculated the chi-square statistics and the degrees of freedom for the implied model (from the goodness-of-fit test). Table 2.9 contains chi-square statistics and degrees of freedom of null and three implied (candidate) models.

Table 2.9
Chi-Square Statistics and Degrees of Freedom of Null and Implied Models

	Null Model	2-Factor Model	3-Factor Model	4-Factor Model
χ^2 (chi-square)	8053.806	761.683	470.022	172.123
df	66	43	33	24

Next, the chi-square statistics and the degrees of freedom were used to calculate absolute and incremental close-fit indexes for all three candidate models. Table 2.10 contains the results of these calculations.

Table 2.10
Calculated Values of Close-Fit Indexes for All Three Models

	Absolute Close-Fit	Incremental Close-Fit		
Model	RMSEA	NFI	CFI	TLI
Two-Factor	.100	.91	.91	.86
Three-Factor	.089	.94	.95	.89
Four-Factor	.061	.98	.98	.95

As these results showed, the two-factor model failed to produce acceptable fit index values (RMSEA was too high and incremental close-fit indexes were too low), whereas both the three-factor and the four-factor models had indexes within or close to their appropriate range/threshold values.

Next, all three candidate models were evaluated based on how well they were aligned with the research constructs. All models correctly placed the environmental attitude items into a separate factor. However, the models produced less than perfect match for the remaining items measuring motivational and science identity constructs. The two-factor model bundled all these items together into a single factor. Even though the four-factor model produced one factor per construct, it placed several items in factors different from their predicted locations. For example, Items 3 and 5 that were placed into the “Science Identity” factor even though they were motivational items. The three-factor model was able to correctly place two values of science items into their own factor. The remaining items were placed into a single factor, therefore, combining science identity and motivational items into one factor. Thus, even though no model produced a perfect match to the predicted factor structure, the three-factor model did a better job than other two models.

Lastly, factor loadings from Table 2.7 showed some uncertainty in choosing of the factors for some items. In the three-factor model, Item # 13 had comparable factor loadings for all three factors (.409, .373, and .437). It was decided to pick factor # 3 for this item since it was a science value item and it made sense to combine it with the other science value item (Item # 11), which had a high factor loading (.751) on that factor. The four-factor model had problems with two items: Item # 3 and Item # 11. Item # 3 had factor loadings of .369, .449, .216, and .261 indicating possibility of placing this item into either factor 1 or factor 2. Likewise, Item # 11 had comparable loadings on factors 1 and 4 (.300 and .494, respectively). Once again, the three-factor model appeared as a better choice than the four-factor model.

Based on the above-mentioned analyses and discussion, it made a perfect sense to choose the three-factor model as the final model and use it in subsequent studies. Indeed, the three-factor model had an acceptable fit to data and aligned well with the research constructs. Figure 2.8 depicts the main steps of exploratory factor analysis of SIEVEA.

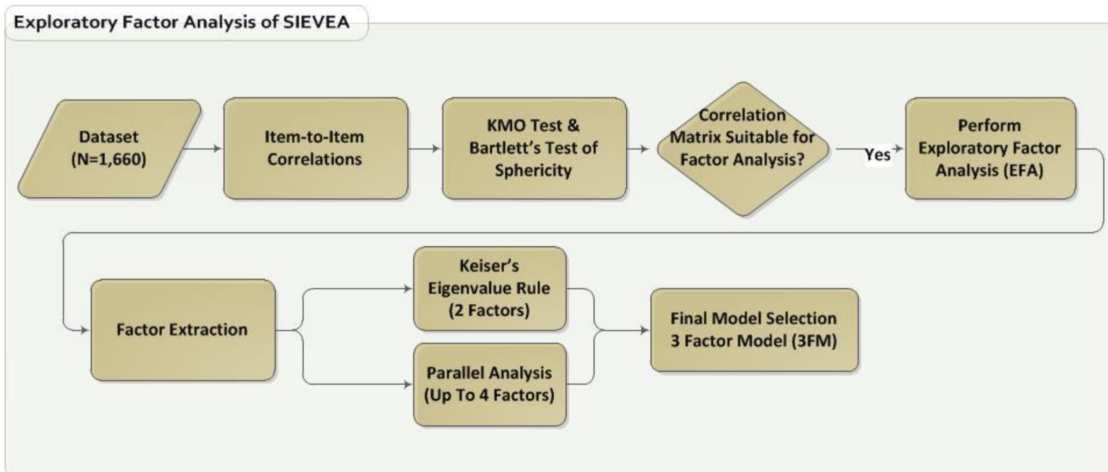


Figure 2.8. Exploratory factor analysis of SIEVEA

Discussion. All three candidate models generated by EFA had proper factor structures that aligned reasonably well with the research constructs. The environmental attitude factor was especially well pronounced in all models.

In the two-factor model items developed for measuring two motivational constructs (expectation of success in science and values of science) and science identity construct were merged into a single factor which was appropriately named "Science Learning Motivation and Self-Efficacy." This result was not surprising since these three construct are closely related. Consequently, in this model four researched constructs were split into 2 factors: "Science Learning Motivation and Self-Efficacy" and "Environmental Attitudes." Hence, this model is suitable when the researcher or the educator is more interested in science learning in general, rather than in any of the specific areas like science identities or motivation.

The four-factor model retained the environmental attitude factor while allowing the splitting of the science learning related factor into its three constituent components. However, the EFA results indicated some mismatch between where items were predicted to cluster versus where they ended up landing. Items 3 and 5 that were originally predicted to belong to Values of Science construct went to factor C1 (Science Identity). Likewise, Item 7 was predicted to belong to Science Identity construct. However, it was placed in factor C2 (Expectations of Success in Science). The remaining items were placed into factors according to their predicted locations based on the previous research and theories. This model can be used when the researcher or the educator is interested in a specific area of science learning, like students' expectations of success in science for example.

The three-factor model was a transitional model from the two-factor model to the four-factor model. It preserved the environmental attitude factor from the two-factor model, but divided the other factor into two factors: science identities and motivation and science values. This model had the cleanest factor structure out of three candidate models.

The final step of EFA was the model selection. For this selection, all candidate models were evaluated based on the following three factors: 1) models' fit to data, 2) how well did each model align with the research constructs, and 3) models' factor loadings. This evaluation showed that the two-factor model did not have a good fit to data. Likewise, even though the four-factor model had acceptable close-fit indexes, its factor structure had some issues and did not align as well with the research constructs as the three-factor model. The three-factor model had the best factor structure out of all models.

Additionally, it had a good fit to data. Therefore, it was decided to pick the three-factor model as the final selection. Only this model will be used for confirmatory factor analysis and validation.

Item #9, “I can use technology for learning science content,” did not fit well within the remaining survey items: its correlation coefficients with other items were noticeably smaller than the in-between coefficients of other items. Additionally, its data had a significant non-normality. Therefore, it was decided to drop this item from the SIEVEA survey instrument and eliminate its data from following analyses. Likewise, this item’s data points will be removed from the collected data prior to performing any additional statistical tests like t-test or ANOVA.

The results of this study illustrated the usefulness of SIEVEA as a simple and expedient instrument for measuring important constructs related to science learning and environmental attitudes. Indeed, the online survey format made this instrument readily available for multiple schools, resulting in extensive student participation.

The SIEVEA instrument has several valuable and convenient features. First, students can take the survey either from home or in the classroom. Second, according to the Lexile framework, the survey’s questions are age appropriate for high school students’ reading level. Third, it takes less than 15 minutes to complete the survey. Next, the survey is short and has only 15 items. Lastly, the survey is online, which makes it very convenient for response data extraction and analysis.

The survey can be used by researchers who study students’ science identities, expectations of success in science, values of science and environmental attitudes. Indeed, this survey is a simple and convenient instrument to measure the researched constructs.

Therefore, the researchers can use it prior and/or post their research related to science learning and environmental attitudes to measure the effectiveness of their researched methods on aforementioned constructs.

The instrument can also be used to survey students' science and environmental attitudes in general. The data generated by this survey can be used by the school administration and teacher to find areas of concern and make changes in instruction and teaching strategies. For example, if the instrument indicates problems with students' motivation in science learning (based on low scores on expectancy of success and science value factors), then specific recommendations can be made in this area of concern.

It took many years to develop motivational theories and draw distinctions between motivational constructs, but what is the practical value of these theories? After all, as Graham and Weiner (1996) argued, people constructed bridges in the past even before engineering courses existed and "healers" saved people's lives before medical schools were opened. However, Graham and Weiner continued, only advances in physics made the construction of the Golden Gate Bridge possible and understanding of biological principles helped to eradicate smallpox. Therefore, it follows that although it is up to an individual to complete or avoid a task, the applications of motivational theories can in fact enhance human performance and because of this, motivational theories should be integrated in education to cultivate classroom motivation (Graham & Weiner, 1996). Table 2.11 adapted from Pintrich (2003) provides an excellent summarization of some of the best design principles for constructing classrooms that promote student interest and motivation.

Table 2.11
Motivational Generalizations and Design Principles (Pintrich, 2003)

Motivational generalization	Design principles
Adaptive self-efficacy and competence beliefs motivate students.	Provide clear and accurate feedback regarding competence and self-efficacy, focusing on the development of competence, expertise, and skill. Design tasks that offer opportunities to be successful but also challenge students.
Adaptive attributions and control beliefs motivate students.	Provide feedback that stresses process nature of learning, including importance of effort, strategies, and potential self-control of learning. Provide opportunities to exercise some choice and control. Build supportive and caring personal relationships in the community of learners in the classroom.
Higher levels of interest and intrinsic motivation motivate students.	Provide stimulating and interesting tasks, activities, and materials, including some novelty and variety in tasks and activities. Provide content material and tasks that are personally meaningful and interesting to students. Display and model interest and involvement in the content and activities.
Higher levels of value motivate students.	Provide tasks, material, and activities that are relevant and useful to students, allowing for some personal identification with school. Classroom discourse should focus on importance and utility of content and activities.
Goals motivate and direct students.	Use organizational and management structures that encourage personal and social responsibility and provide a safe, comfortable, and predictable environment. Use cooperative and collaborative groups to allow for opportunities to attain both social and academic goals. Classroom discourse should focus on mastery, learning, and understanding course and lesson content. Use task, reward, and evaluation structures that promote mastery, learning, effort, progress, and self-improvement standards and less reliance on social comparison or norm-referenced standards.

To summarize, the SIEVEA is a valuable tool for analyzing high school student's science identities, motivational and environmental attitudes. However, the study had several limitations that should be noted.

The study included both urban and suburban schools. All schools were public schools except for one, which was a private school. Although this selection of schools should have helped in making the sample socioeconomically diverse, the exact demographics of the study are not known, other than gender. The anonymous nature of the study was extremely helpful in getting more school districts to participate. However, a considerable disadvantage of this was that it restricted the generalizability of the results of the study.

The second limitation of the study was that students' grade level information was not collected. Because of this, the results present a combination of all high school grade levels. It may be useful to collect grade level data in future studies so results can be validated for specific high school grades. The fact that two survey questions were self-developed rather than adapted from prior studies may also be considered as a limitation of this study (Stone, 1978).

Conclusion. When researchers are interested in understanding what motivates students to learn science, they usually examine their reasons of science learning and their beliefs and feelings that influence this learning (Glynn et al., 2008). It is believed that understanding the factors that contribute to students' motivation of science learning will help science researchers and educators to improve science education.

For example, various studies of theoretical models of motivation pointed out that when students hold strong beliefs and expectations about themselves, they are more likely to engage in tasks and persist in doing them despite any difficulties they may encounter while working on tasks (Weiner, 1992). Likewise, understanding students' attitudes toward environment can help environmental educators to take actions for increasing students' environmental awareness and their environmental literacy.

The survey developed in this study can help researchers and educators in accomplishing the above-mentioned tasks. The SIEVEA measures high school students' science identities, expectations of success in science, values of science and environmental attitudes. Its predicted usefulness was confirmed by analyzing, refining and interpreting the collected data. The data analysis showed the instrument's usefulness in measuring the researched constructs. Future research should be conducted to confirm the relationships between SIEVEA factor scores and students' motivational and environmental attitude behaviors.

The instrument can serve as a screening/diagnostic tool for science teachers who want to identify students who lack motivation or have indifferent attitudes toward the environment. On the other hand, it can be used as an instrument for evaluating newly introduced science curriculum or instructional methodology. In addition to being an appropriate tool for measuring the aforementioned constructs, it is also convenient and easy to use. Tyler-Wood, Knezek & Christensen (2010) stressed that to be effective, the instrument should be short, straightforward to administer and easy to understand.

References

Agarwal, N. K. (2011). Verifying survey items for construct validity: A two stage sorting procedure for questionnaire design in information behavior research. *Proceedings of the American Society for Information Science and Technology, 48*(1), 1-8.

Andre, T., Whigham, M., Hendrickson, A., & Chambers, S. (1999). Competency beliefs, positive affect, and gender stereotypes of elementary students and their parents about science versus other school subjects. *Journal of Research in Science Teaching, 36*(6), 719-747.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Macmillan.

Baumgartner, H., & Hombur, C. (1996). Applications of structural equation modeling in marketing and consumer research: A review. *International Journal of Research in Marketing, 13*, 139-161.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*, 238-246.

Bentler, P. M. (1995). *EQS structural equations program manual.* Encino, CA: Multivariate Software.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*, 588-606.

Blatt, E. (2014). Uncovering students' environmental identity: An exploration of activities in an environmental science course. *The Journal of Environmental Education*, *45*(3), 194-216.

Bollen, K. A. (1989). A new incremental fit index for general structural equation models. *Sociological Methods & Research, 17*(3), 303-316.

Bradley, J. C., Waliczek, T. M., & Zajicek, J. M. (1999). Relationship between environmental knowledge and environmental attitude of high school students. *The Journal of Environmental Education, 30*(3), 17-21.

Browne, M. W. (1990). *MUTMUM PC user's guide*. Columbus: Ohio State University, Department of Psychology.

Browne, M. W., & Cudek, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing Structural Equation Models* (pp. 136–162). Newbury Park, CA: Sage.

Byrne, B. M. (2013). *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. New York, NY: Routledge.

Carlone, H. B., & Johnson, A. (2007). Understanding the science experiences of successful women of color: Science identity as an analytic lens. *Journal of Research in Science Teaching, 44*(8), 1187-1218.

Child, D. (2006). *The essentials of factor analysis*. (3rd ed.). New York, NY: Continuum International Publishing Group.

Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis*. Hillsdale, NJ: Erlbaum.

Costello, A. B., & Osborne, J. W. (2011). *Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis*. Retrieved from http://pareonline.net/pdf/v10n7.pdf

DeVellis, R. (2003). *Scale development: Theory and applications.* Thousand Oaks, CA: Sage.

DeVellis, R. F. (2012). *Scale development: Theory and applications* (3rd ed.). Los Angeles, CA: Sage.

Dunlap, R. E., & Van Liere, K. D. (1978). The “new environmental paradigm”. *The Journal of Environmental Education, 9*(4), 10-19.

Dunlap, R. E., Van Liere, K. D., Mertig, A. G., & Jones, R. E. (2000). New trends in measuring environmental attitudes: measuring endorsement of the new ecological paradigm: a revised NEP scale. *Journal of Social Issues*, *56*(3), 425-442.

Eccles, J. S., & Wigfield, A. (1995). In the mind of the actor: The structure of adolescents' achievement task values and expectancy-related beliefs. *Personality and Social Psychology Bulletin*, *21*(3), 215-225.

Eccles, J. S., Wigfield, A., & Schiefele, U. (1998). Motivation to succeed. In W. Damon (Series Ed.) & N. Eisenberg (Vol. Ed.), *Handbook of Child Psychology: Vol. 3. Social, emotional and personality development* (5th ed., pp. 1017–1094). New York, NY: Wiley.

Eilam, E., & Trop, T. (2012). Environmental attitudes and environmental behavior—Which is the horse and which is the cart? *Sustainability*, *4*(9), 2210-2246.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*(3), 272-299.

Field, A. (2005). *Discovering statistics with SPSS* (2nd Edition). London, UK: Sage.

Falk, J. H. (2009). *Identity and the museum visitor experience*. Walnut Creek, CA: Left Coast Press.

Fortus, D., & Vedder-Weiss, D. (2014). Measuring students' continuing motivation for science learning. *Journal of Research in Science Teaching*, *51*(4), 497-522.

Gehlbach, H., & Brinkworth, M. E. (2011). Measure twice, cut down error: A process for enhancing the validity of survey scales. *Review of General Psychology*, *15*(4), 380-387.

Gignac, G. E. (2009). Partial confirmatory factor analysis: Described and illustrated on the NEO–PI–R. *Journal of Personality Assessment*, *91*(1), 40-47.

Glynn, S. M., Brickman, P., Armstrong, N., & Taasoobshirazi, G. (2011). Science motivation questionnaire II: Validation with science majors and nonscience majors. *Journal of Research in Science Teaching*, *48*(10), 1159-1176.

Glynn, S. M., & Koballa, T. R. (2006). Motivation to learn in college science. In J. J. Mintzes & W. H. Leonard (Eds.), *Handbook of college science teaching* (pp. 25-32). Arlington, VA: National Science Teachers Association Press.

Glorfeld, L.W. (1995). An improvement on Horn's parallel analysis methodology for selecting the correct number of factors to retain. *Educational and Psychological Measurement*, *55*, 377-393.

Golafshani, N. (2003). Understanding reliability and validity in qualitative research. *The Qualitative Report*, *8*(4), 597-606.

Gorsuch, R. L. (2003). Factor analysis. In J. A. Schinka & F. Velicer (Eds.), *Handbook of psychology: Vol. 2. Research Methods in Psychology* (pp. 143-164). Hoboken, NJ: John Wiley.

Graham, S., & Weiner, B. (1996). Theories and principles of motivation. *Handbook of Educational Psychology*, *4*, 63-84.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*, 179-185.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1-55.

Hungerford, H. R., Litherland, R. A., Volk, T. L., Ramsey, J. M., & Peyton, R. B. (2003). *Investigating and evaluating environmental issues and actions: Skill development program*. Champagin, IL: Stipes Publishing, LLC.

Johnson, B., & Manoli, C. C. (2011). The 2-MEV scale in the United States: A measure of children's environmental attitudes based on the theory of ecological attitude. *The Journal of Environmental Education, 42*(2), 84-97.

Joppe, M. (2000). *The research process.* Retrieved from http://www.htm.uoguelph.ca/pagefiles/MJResearch/ResearchProcess/home.html.

Kaiser, H.F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement, 20,* 141-151.

Kline, P. (2014). *An easy guide to factor analysis.* New York, NY: Routledge.

Kline, R. B. (2005). *Principles and Practice of Structural Equation Modeling*, (2nd ed.). New York: Guilford.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, *140*, 5-55.

Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin*, *103*(3), 391.

Matsui, T., Matsui, K., & Ohnishi, R. (1990). Mechanisms underlying math self-efficacy learning of college students. *Journal of Vocational Behavior*, *37*(2), 225-238.

Matsunaga, M. (2010). How to factor-analyze your data right: Do's, don'ts, and how-to's. *International Journal of Psychological Research*, *3*(1), 97-110.

MetaMetrics, Inc. (2012). *The lexile framework for reading*. Retrieved from www.lexile.com.

Mubeen, S., & Reid, N. (2014). The measurement of motivation with science students. *European Journal of Educational Research, 3*(3), 129-144.

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

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instrumentation, and Computers, 32*(3), 396-402.

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.

Plante, I., O'Keefe, P. A., & Théorêt, M. (2013). The relation between achievement goal and expectancy-value theories in predicting achievement-related outcomes: A test of four theoretical conceptions. *Motivation and Emotion*, *37*(1), 65-78.

Pintrich, P. R. (2003). A motivational science perspective on the role of student motivation in learning and teaching contexts. *Journal of Educational Psychology*, *95*(4), 667-686.

Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, *82*(1), 33-40.

Rea, L. M., & Parker, R. A. (2014). *Designing and conducting survey research: A comprehensive guide*. San Francisco, CA: John Wiley & Sons.

Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. *Psychological Bulletin*, *114*(3), 552.

Russell, D. W. (2002). In search of underlying dimensions: The use (and abuse) of factor analysis in Personality and Social Psychology Bulletin. *Personality and Social Psychology Bulletin*, *28*(12), 1629-1646.

Sadler, T.D. (2009). Situated learning in science education: Socio-scientific issues as contexts for practice. *Studies in Science Education*, *45*, 1-42.

Steiger, J. H., & Lind, J. C. (1980, May). *Statistically based tests for the number of common factors.* Paper presented at the annual meeting of the Psychometric Society, Iowa City, IA.

Stetsenko, A. (2008). From relational ontology to transformative activist stance on development 153 and learning: Expanding Vygotsky's (CHAT) project. *Cultural Studies of Science Education*, *3*, 471-491.

Stone, E. F. (1978). *Research methods in organizational behavior.* Santa Monica, CA: Goodyear Publishing Company.

Swarat, S., Ortony, A., & Revelle, W. (2012). Activity matters: Understanding student interest in school science. *Journal of Research in Science Teaching*, *49*(4), 515-537.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics.* Boston, MA: Allyn and Bacon.

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

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, *38*, 1-10.

Tyler-Wood, T., Knezek, G., & Christensen, R. (2010). Instruments for assessing interest in STEM content and careers. *Journal of Technology and Teacher Education, 18*(2), 345-368.

Warner, J. M. (2008). Supplemental instruction for non-science majors biology students: Meanings and influences on science identities for women. *Dissertation Abstracts International*, *69*(7), 2658A.

Weiner, B. (1992). *Human motivation: Metaphors, theories, and research*. Newbury Park, CA: Sage Publications.

Zecha, S. (2010). Environmental knowledge, attitudes and actions of Bavarian (southern Germany) and Asturian (northern Spain) adolescents. *International Research in Geographical and Environmental Education, 19*(3), 227-240.

Zwick, W. R. & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin, 99*, 432-442.

Appendix A

Survey: SIEVEA

1. What is your gender?
    - Female
    - Male
2. My favorite science subject is (pick one)
    - Biology
    - Chemistry
    - Earth Science
    - Environmental Science
    - Forensics
    - Physics
    - None
    - Other
3. Learning science in school will help me to succeed later in life.
    - Strongly Agree
    - Agree
    - Neither Agree nor Disagree
    - Disagree
    - Strongly Disagree
4. I am confident I can master the skills taught in my science class.
    - Strongly Agree
    - Agree

- Neither Agree nor Disagree
- Disagree
- Strongly Disagree

5. I consider science topics very interesting and engaging.
    - Strongly Agree
    - Agree
    - Neither Agree nor Disagree
    - Disagree
    - Strongly Disagree
6. When it comes to learning science, I think of myself as a science person.
    - Strongly Agree
    - Agree
    - Neither Agree nor Disagree
    - Disagree
    - Strongly Disagree
7. My peers and teachers think that I am knowledgeable in science.
    - Strongly Agree
    - Agree
    - Neither Agree nor Disagree
    - Disagree
    - Strongly Disagree
8. I am certain I can figure out how to do the most difficult science class work.
    - Strongly Agree

- Agree
- Neither Agree nor Disagree
- Disagree
- Strongly Disagree

9. I can use technology for learning science content.
   - Strongly Agree
   - Agree
   - Neither Agree nor Disagree
   - Disagree
   - Strongly Disagree
10. My friends and family recognize me as a scientist.
    - Strongly Agree
    - Agree
    - Neither Agree nor Disagree
    - Disagree
    - Strongly Disagree
11. It is important to me that I look smart in my science class.
    - Strongly Agree
    - Agree
    - Neither Agree nor Disagree
    - Disagree
    - Strongly Disagree
12. I would like to become more active on important environmental issues.

- Strongly Agree
- Agree
- Neither Agree nor Disagree
- Disagree
- Strongly Disagree

13. One of my goals is to show others that I am good at science.

- Strongly Agree
- Agree
- Neither Agree nor Disagree
- Disagree
- Strongly Disagree

14. It is important for all people to be engaged in vital environmental issues.

- Strongly Agree
- Agree
- Neither Agree nor Disagree
- Disagree
- Strongly Disagree

15. I am interested in reading websites, articles or watching TV programs, documentary movies about the environmental issues.

- Strongly Agree
- Agree
- Neither Agree nor Disagree
- Disagree

- ○ Strongly Disagree

Study 1B – Validation of SIEVEA via a Confirmatory Factor Analysis

Introduction. Factor analysis is a set of popular statistical techniques used by researchers who are interested in investigating the variability and relationships between observed variables and unobserved, latent constructs. Yet, there is strong evidence that many researchers fail to conduct factor analysis properly (Fabrigar, Wegener, MacCallum, & Strahan, 1999; Hanson & Roberts, 2006). In order to avoid common pitfalls and conduct a robust factor analysis of data collected using the SIEVEA survey instrument, this study followed the strict guidance of Gignac (2009), Jackson, Gillapsy & Purc-Stephenson (2009), Matsunaga (2010) and Russell (2002). Before proceeding with this study's aims and results, it is useful to discuss what factor analysis is, what different types of factor analysis are available and how they fit in the bigger picture.

As noted before, factor analysis is comprised of commonly used statistical methods for exploring variability and relationships of the collected data in order to identify and confirm the underlying constructs and how they are mapped to the observed variables. Simply put, factor analysis allows for analysis of the data and confirmation of whether the variables used really measured the constructs they were intended to measure (Matsunaga, 2010).

There are two main methods of factor analysis: exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). These methods serve different purposes in the overall process of factor analysis. The EFA's role is to provide a glimpse into the data's underlying factor structure and to come up with a plausible theoretical model. On the other hand, the CFA works with a priori model produced by the EFA and tests its viability or fit by comparing proposed factor structure to data. The CFA confirms the model’s fit by calculating and examining various model fit indexes (Bandalos, 1996; Russel, 2002). To summarize, “exploratory factor analysis helps to define the internal structure for a set of items and to group the items into factors,” whereas “confirmatory factor analysis needs to be used to establish construct validity” (Pett, Lackey, & Sullivan, 2003, p. 239). Hence, the EFA is used for theory building, while the CFA is a theory-testing tool (Matsunaga, 2010). Furthermore, the CFA is a useful analytical tool for evaluating instruments’ construct validity and invariance measurement across groups (Brown, 2015), as well as for validating measurement models (MacCallum & Austin, 2000).

This study is a continuation of prior research. In the prior study, the SIEVEA survey instrument was developed to measure high school students’ science identities, expectation of success, values of science and environmental attitudes. The survey was used to collect a large sample (1,764 responses), which was subsequently analyzed using exploratory factor analysis. The EFA of the previous study provided useful insights into the factor structure of the data and led to the formation of plausible models. Three candidate models were suggested: the two-factor, the three-factor and the four-factor models. Further analyses of three candidate models along with their close-fit indexes and factor loadings led to the selection of the three-factor model as the final model.

The purpose of the current research is to conduct a CFA in order to evaluate the three-factor model and confirm or reject its factor structure produced by the EFA. Additionally, confirming the factor structure will further enhance the SIEVEA instrument’s construct validity. This is important, because the EFA conducted in the previous study did not fully establish the instrument’s construct validity.

Partial-confirmatory factor analysis. Before continuing with confirmatory factor analysis, an intermediate step, called partial-confirmatory factor analysis (PCFA), was implemented. Gignac (2009) noticed that many researchers omit this important step and insisted that a PCFA should precede confirmatory factor analysis (CFA) to confirm the model was derived from exploratory factor analysis (EFA). Indeed, most CFAs are not strictly confirmatory, but rather contain some elements of ambiguity and end up modifying the models in order to reach an appropriate model-fit (Nesselroade, 1994). Therefore, a PCFA is a useful intermediate step between EFA and CFA and can be used to strengthen the model's plausibility and justify the necessity of conducting a CFA. This approach is similar to implementation of the Kaiser-Mayer-Olkin (KMO) index and Bartlett's sphericity tests to substantiate the case of conducting exploratory factor analysis.

Method. *Sample.* The sample data was made up of survey responses of 1,764 high school students, out of which 52.72% were females and 46.88% were males. 0.4% of students did not report their gender. This was the exact same sample that was used for conducting an EFA. Data was collected from 11 urban, suburban and private high schools from New Jersey, Pennsylvania and Connecticut.

Measure. The SIEVEA instrument was developed for measuring the following four constructs: science identities, expectations of success, values of science and environmental attitudes. The collected data was analyzed using EFA. When determining the extraction of an appropriate number of factors, Kaiser's (1960) eigenvalue-greater-than-one rule suggested two factors.

However, Parallel Analysis (PA), which is considered the most accurate method for determining the number of factors to be retained (Reise, Waller, & Comrey, 2000), indicated an extraction of up to four factors. Additional analyses led to the selection of the three-factor model. This model was subjected to PCFA in order to determine its fit to data prior to performing a CFA on it.

Data preparation. Before conducting PCFA, responses with one or more missing data points were excluded. This approach to taking into account missing data is known as listwise deletion. It is the most commonly used method for dealing with missing data (Schafer & Graham, 2002) and is the default option in many statistical packages. Listwise deletion resulted in removal of 104 responses (5.9% of total). The remaining data set had 1,660 responses out of the initial 1,764 responses.

Data analysis. In performing PCFA, this study closely followed Gignac's (2009) guidelines. According to Gignac, a PCFA is made up of the following steps: 1) specify a meaningful candidate model (also called an implied model) based on prior analysis like EFA, 2) estimate the model via maximum likelihood estimation (MLE), and 3) evaluate the model's appropriateness by calculating suitable statistics and close-fit indexes and comparing them to established cutoff values (Gignac, 2009).

The first step was satisfied during the prior EFA study, which suggested the three-factor model as a candidate (implied) model. The next step was to perform a factor analysis using maximum likelihood estimation, which "… will yield a chi-square value (and degrees of freedom) associated with the residual correlation matrix" (Gignac, 2009, p. 41). SPSS (Version 24) statistical package provides an extraction method called 'maximum likelihood' which can be used for this analysis and produces the goodness-of-fit chi-square statistic and the degrees of freedom of the implied model.

The MLE method, along with varimax (orthogonal) rotation, was applied to extract 3 (for three-factor model) factors from the 12x12 correlation matrix. Furthermore, choosing Bartlett's test of sphericity option calculated the chi-square statistic and the degrees of freedom of the null model. The resulting values of chi-square statistics (from the goodness-of-fit test and the Bartlett's test of sphericity) and the degrees of freedom for the null and the implied model are listed in Table 2.12.

Table 2.12
Chi-Squares and Degrees of Freedom of Null and Three-Factor Models

	Null Model	3-Factor Model
χ^2 (chi-square)	8053.806	470.022
df	66	33

Model-fit evaluation. The model's fit evaluation during PCFA and CFA is done by computing suitable absolute and incremental close-fit statistics of the model and comparing them to known cutoff values (Gignac, 2009). There is no shortage of absolute and incremental close-fit indexes used for model-fit evaluation (Baumgartner & Hombur, 1996; Bollen, 1989). In order to evaluate the three-factor model during PCFA and CFA, this study utilized the following commonly used close-fit indexes: *the Root Mean Square Error of Approximation* (RMSEA; Browne & Cudeck, 1993), *the Standardized Root Mean Residual* (SRMR; Bentler, 1995), *the Normal Fit Index* (NFI; Bentler & Bonnett, 1980), *the Tucker-Lewis Index* (TLI; Tucker & Lewis, 1973), and *the Comparative Fit Index* (CFI; Bentler, 1990).

Since each of these indexes evaluates the model from a different perspective (Kline, 2005), it made perfect sense to examine as many of them as possible during PCFA and CFA in order to make accurate conclusions about the fit of the proposed model.

These close-fit indexes were defined and thoroughly discussed in the prior study (Study 1A). Four indexes, RMSEA, NFI, TLI and CFI, are straightforward to compute since their calculations use fairly simple formulas (Gignac, 2009). These computations are based on the chi-square statistic and the degrees of freedom of the null and implied models which were calculated above. Unfortunately, the manual calculation of SRMR is not easy: it is based on the residual correlation matrix produced by factor analysis and is computation intensive (Bentler, 1995). Therefore, in order to calculate SRMR, it is required to use the statistical packages like SPSS.

The cutoff values of close-fit indexes were also covered in Study 1A. To summarize, for absolute close-fit indexes like RMSEA and SRMR, values of .08 or less suggest an acceptable model fit (Browne, 1990; Gignac, 2009; Hu & Bentler, 1999), whereas values greater than .10 indicate a poor fit (Browne & Cudeck, 1993). The incremental close-fit indexes like NFI, TLI and CFI, have different thresholds for acceptable and good model fits. An index value of .95 or higher is considered as a good fit (Hu & Bentler, 1999; Thompson, 2004), whereas the cutoff value for an acceptable fit is .90 (Bentler & Bonett, 1980; Reise, Widaman, & Pugh, 1993; Russell, 2002).

Results. The final step of PCFA was to compute close-fit indexes and compare them to established cutoff values (Gignac, 2009) in order to decide on model's suitability for CFA. The chi-square statistics of the null and the implied model, along with their degrees of freedom (see Table 2.12), were used to calculate absolute and incremental close-fit indexes. One absolute close-fit index (RMSEA) and three incremental close-fit indexes (NFI, CFI, and TLI) were calculated for the three-factor model. These calculations resulted in the following close-fit indexes values: RMSEA = .089, NFI = 0.94, CFI = .95, TLI = .89.

As these results showed, the three-factor model had indexes within or close to their appropriate range/threshold values. Therefore, the three-factor model passed PCFA and could be subjected to CFA. Figure 2.9 displays the main steps of PCFA of SIEVEA.

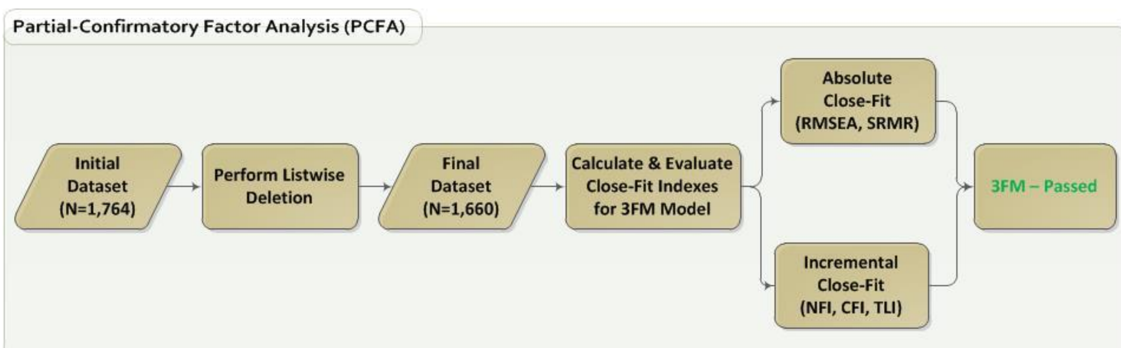


Figure 2.9. Partial-confirmatory factor analysis of SIEVEA

Confirmatory factor analysis. CFA is conducted when the models are already established from the theory and then need to be tested (Eccles & Wigfield, 1995). CFA allows the researchers to test their initial theory about the constructs present in data; therefore, one should have a solid theory regarding the number of latent variables in a model (Thompson, 2004).

Typically, CFA is carried out using the instrument's individual items. However, occasionally, the researchers use composite scores, also known as parcels, as an alternative for individual items (Usher & Pajares, 2009). These composite scores are either the sum or average of a group of individual items. They are used as observed variables in structural equation models (SEM) and in CFA (Little, Cunningham, Shahar, & Widaman, 2002).

For example, Lent, Lopez, Brown, & Gore (1996) performed confirmatory factor analysis using parcels on multiple (four) candidate model out of which only one model, with five factors, was accepted as the best fit to data. Other researchers used parcels during CFA for the decision of whether a single latent variable was supported by the parceled scores (Stevens, Olivarez, & Hamman, 2006). Despite the preceding, in addition to other examples of parcel usage, there is no consensus among researchers regarding the use of item parcels. More empirical research is needed for identifying the proper use cases of the parceling method, as well as the situations where it should be avoided (Little, Rhemtulla, Gibson, & Schoemann, 2013). Currently, the parceling techniques are most commonly used for improving the sample size to model size ratio (Williams & O'Boyle, 2008). As the sample size of this study was large and the model size was small, it made sense to avoid parcels and proceed with performing CFA on the original individual items.

Before executing a CFA, the model needed to be specified. As PCFA results showed, the three-factor model had acceptable fit values and, therefore, was a good candidate to pass CFA. Consequently, this model was tested using a new sample data for conformity of its suggested factor structure to the data.

Method. *Sample.* The sample data was made up of survey responses of 1,495 high school students out of which 814 (54.44%) were females and 677 (45.28 %) males. Four students (0.28 %) did not report their gender. Data was collected from three schools (urban and suburban) in New Jersey and Connecticut. This data was specifically gathered for conducting a CFA.

Prerequisites. PCFA performed on the three-factor model prior to attempting a CFA produced the following fit index values: RMSEA = .089, NFI = 0.94, CFI = .95, TLI = .89. These results indicated the suitability of performing CFA.

Data analysis. Although many goodness-of-fit indexes are available for evaluating candidate models in CFA, this study used the most popular and frequently used goodness-of-fit indexes. Since the sample size of the current study was large, it was expected that the chi-square value would be significant. Therefore, to avoid common mistakes made by many researchers who conducted CFA (Jackson, Gillaspy, & Purc-Stephenson, 2009), close-fit indexes that are less sensitive to the sample size were calculated. Based on Matsunaga's (2010) recommendations, the following absolute and incremental close-fit indexes were computed: RMSEA, SRMR, NFI, CFI, and TLI.

Evaluation of the model. CFA was executed on the three-factor model using Analysis of Moment Structures (AMOS) software (Version 24). AMOS is a popular tool for carrying out structural equation modeling (SEM). It provides different evaluation methods that can be used to evaluate models. In this study, the maximum likelihood (ML) method, the most common model estimation procedure for conducting CFA, was implemented.

The study closely followed the guidelines and procedures recommended by Matsunaga (2010) and Jackson, Gillaspy, & Purc-Stephenson (2009). This helped in avoiding common mistakes that happen during factor analysis. While testing the models' fit to data and calculating relevant close-fit indexes, special attention was paid to evaluated factor loadings, correlations and their statistical significance. The models were repeatedly adjusted and reevaluated in order to improve the fit.

Data preparation. Prior to performing CFA using ML, it is necessary to deal with missing data. Several approaches are available for handling missing data: listwise deletion, pairwise deletion, substitution of missing values with means of other values, etc. The most common method is listwise deletion (Schafer & Graham, 2002). In listwise deletion, all cases with one or more missing data points are dropped from the analysis. Hence, the analysis is conducted with 100% complete cases. This study used listwise deletion. As a result, 56 incomplete responses out of the total 1,495 responses were dropped. CFA was performed on the remaining data, totaling 1,439 responses.

Normality of data and outliers. The maximum likelihood (ML) estimation procedure assumes that data is multivariate normally distributed (Jackson, Gillaspy, & Purc-Stephenson, 2009). Therefore, prior to conducting CFA using ML, data should be checked for multivariate normality and any deviations from normality should be addressed.

Both Kolmogorov-Smirnov and Shapiro-Wilk tests of normality, using sample data for each item, indicated that there were some deviations from normal distribution. Since "normality on each of the variables separately is a necessary, but not sufficient, condition for multivariate normality to hold" (Stevens, 1996, p. 243), these tests indicated that the data does not fully satisfy multivariate normality assumption of ML. Additionally, AMOS allows for testing of data normality and outliers. Performing this test demonstrated multivariate kurtosis of 25.545, with a critical ratio of 26.433. These values again indicated some non-normality in data.

Several strategies are available for dealing with the non-normality of data. Applying various data transformation, like square-root transformation or logarithm transformation, can substantially improve the distribution of measured variables (Jackson, Gillaspy, & Purc-Stephenson, 2009). Likewise, finding and removing data outliers can reduce non-normality in data. AMOS can detect data outliers defined as observations farthest from the centroid (Mahalanobis distance). After removing some outliers, there was a significant reduction in multivariate kurtosis: it became 6.556, with a critical ratio of 6.544. However, this still indicated considerable non-normality (critical ratio should be 1.96 or less for multivariate normal data). Therefore, different approach from outlier removal was determined for implementation.

Bootstrap resampling, available in AMOS, is another option for working under non-normal data conditions (Byrne, 2013). The bootstrap was proposed by Efron (1979) and was used in many statistical procedures including exploratory and confirmatory factor analysis (Bollen & Stine, 1992; Ichikawa & Konishi, 1995). Bootstrapping allows for use of ML and attainment reliable results even when data normality is clearly violated or the sample is not sufficiently large (Nevitt & Hancock, 2001; Tsagkanos, 2007). Bootstrap works by repeatedly creating random samples from the original data set using sampling with replacement, followed by ML estimation using each sample. Then, the results of these multiple estimations are used to derive the final ML results.

AMOS provides a simple interface for specifying bootstrap ML parameters and running it. Additionally, doing bootstrap does not require any data preparations, like transformation or removal of outliers. Due to these advantages of bootstrap and its ease of use within AMOS software, it was decided to execute CFA using the bootstrap ML option.

The three-factor model. In order to do CFA in AMOS, it is required to set up the model as a path diagram with circles representing the latent concepts and squares representing observed variables. Since this model has three factors and thirteen items, three circles were drawn for three latent variables along with eight, three and two squares for observed variables, each set of squares connected to its respective latent variable as suggested by EFA. One-directional arrows used to connect latent variables to observed variables (survey items), indicated assumed causal influence of latent variables on corresponding items. Double-directional arrows were used to represent covariance among three latent variables.

Additionally, since factor analysis assumes that the observed variables were measured with errors (Matsunaga, 2010), the diagram needed to capture the measurement errors. This was accomplished by adding unobserved variables, one error variable per item (i.e., e3 for Item #3).

Bootstrap ML was executed with Number of Bootstrap Samples parameter set to 2,000. The resulting measurement model for the three-factor model is shown in Figure 2.10.

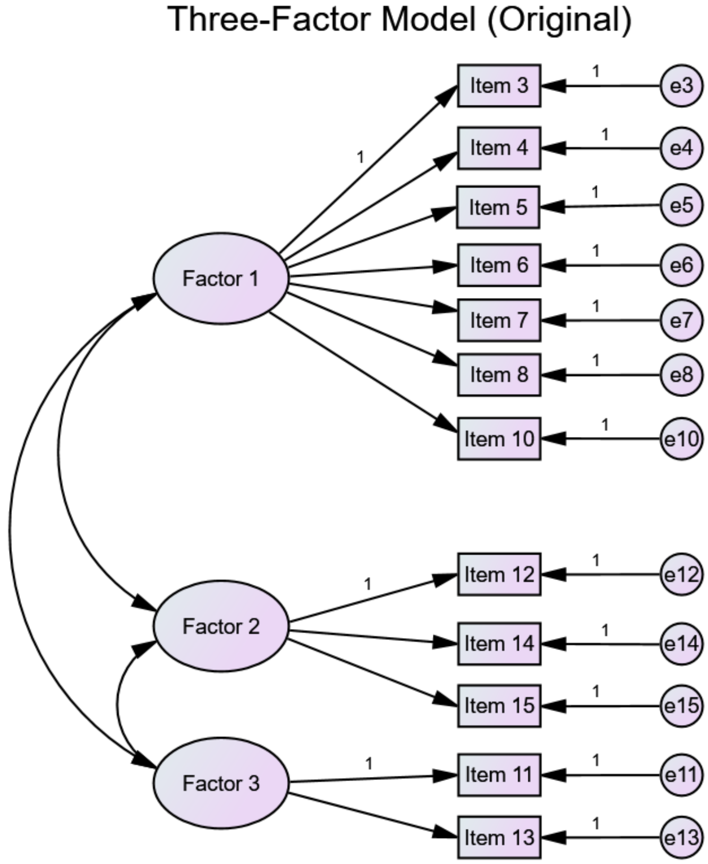


Figure 2.10. The Representation of Three-Factor Model as a Path Diagram. Oval shapes represent unobservable latent factors, rectangles are representations of observed items and arrows indicate loadings.

All standardized factor loadings were significant with $p < .001$ and ranged in magnitude from .608 to .873. As data in Table 2.13 shows, these factor loadings closely resemble those of EFA even though the data set used for CFA was different from the one used for EFA. Inter-factor correlations were .48, .51 and .63, indicating some correlation between factors. This was not surprising because, as Matsunaga (2010) emphasized, most phenomena that are studied in social sciences are interrelated and CFA models usually specify inter-relations among latent variables/factors.

Table 2.13

Unrestricted (EFA) and Restricted (CFA) Three-Factor Model: Standardized Factor Loadings

	Unrestricted (EFA)			Restricted (CFA)		
Item	C1	C2	C3	C1	C2	C3
3	**.605**	.268	.152	**.677**	*	*
4	**.630**	.054	.146	**.634**	*	*
5	**.686**	.352	.023	**.807**	*	*
6	**.794**	.304	.087	**.873**	*	*
7	**.582**	.028	.277	**.636**	*	*
8	**.607**	.063	.201	**.608**	*	*
10	**.650**	.260	.164	**.775**	*	*
11	.246	.197	**.751**	*	*	**.680**
12	.120	**.753**	.148	*	**.799**	*
13	.409	.373	**.437**	*	*	**.828**
14	.104	**.677**	.140	*	**.713**	*
15	.247	**.613**	.054	*	**.619**	*

Note. EFA was performed on the original data set (1,660 responses), whereas CFA was performed on the new data set (1,439 responses).

Additionally, the inspection of Modification Indices for covariances demonstrated that the treating of covariances between some error variables as free parameters, provided the variables were within the same factor, could further improve the model's fit. The final measurement model for the three-factor model resulted in the following close-fit indexes values: RMSEA = .071, SRMR = .044, NFI = 0.95, CFI = .95, TLI = .94. Figure 2.11 depicts the final measurement model for the three-factor model after the above-mentioned changes were applied.

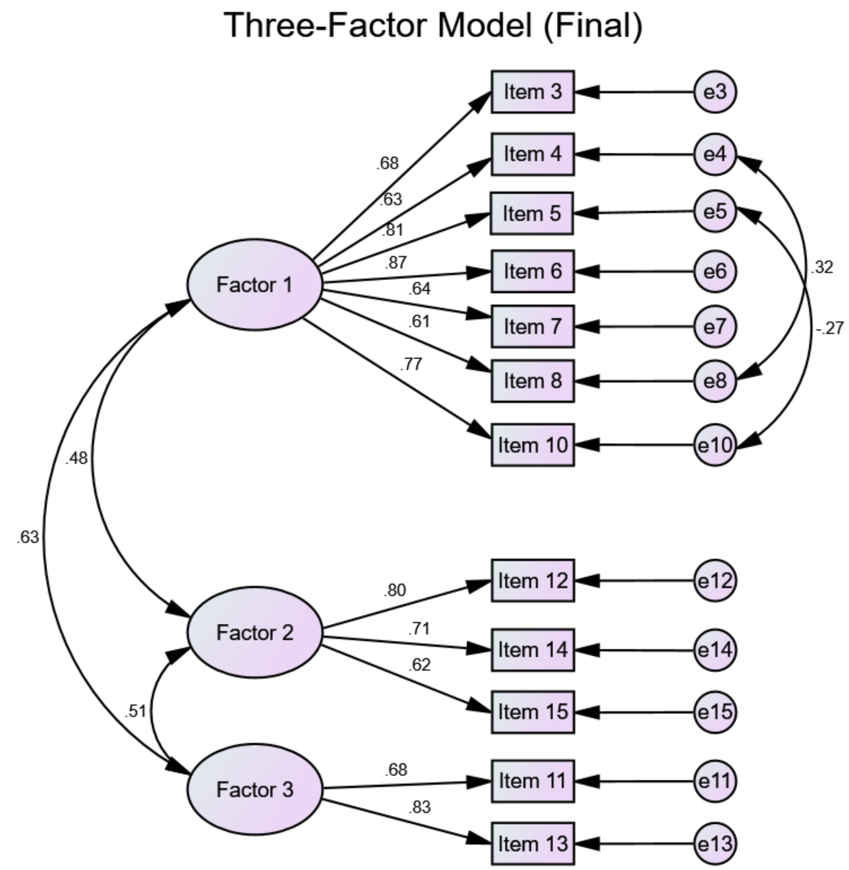


Figure 2.11. The Representation of Final Three-Factor Model
Factor loadings and correlations between factors are displayed. Also shows additional free parameters between error variables: e4 and e8; e5 and e10.

Results. CFA generated values of absolute and incremental close-fit indexes for the three-factor model are summarized in Table 2.14.

Table 2.14
CFA Calculated Close-Fit Indexes for Three-Factor Model

Absolute Close-Fit		Incremental Close-Fit		
RMSEA	SRMR	NFI	CFI	TLI
.071	.044	.95	.95	.94

The close-fit indexes of the three-factor model showed an acceptable fit, with some indexes approaching the thresholds of a good fit. Both the RMSEA (.071) and the SRMR (.044) were below the .08 threshold of the acceptable fit. The incremental close-fit indexes were either equal to or slightly less than the borderline value of .95 (NFI = 0.95, CFI = .95, TLI = .94).

Tests for measurement invariance. When comparing scores between different groups, it is assumed that the instrument (e.g., survey) measures the same psychological constructs in these groups. Furthermore, the comparisons cannot be considered meaningful unless this assumption holds (Milfont & Fischer, 2010). Indeed, if the instrument does not work the same way across different groups, then any comparison between groups using its results will be either biased or unreliable. Therefore, it is very crucial to establish measurement invariance (MI) across different groups (e.g., females vs males) before the instrument's results can be used to meaningfully compare any differences between these groups (Schoot, Lugtig, & Hox, 2012). Measurement invariance is a requirement for a good model, as it validates that the same constructs are being measured across the groups (Hortensius, 2012).

The checklist for testing measurement invariance (Schoot, Lugtig, & Hox, 2012) was used as a guideline for conducting the following analysis of the measurement invariance. In order to test the model for measurement invariance, it is necessary to conduct several multi-group confirmatory factor analyses (MGCFA). These analyses are performed on increasingly restrictive models starting with the theoretical model. During each step, additional constraints, indicating increasingly stringent measurement invariance, are imposed on the model. This leads to a hierarchical ordering of models, with each level having a fewer number of parameters and more degrees of freedom than the prior level. Next, these models are tested using goodness-of-fit indexes and are compared to the prior model using the chi-squared difference test. If the model provides a good fit to the data and the chi-squared difference test indicates an improvement in this model's fit in comparison to the prior model, then the model is accepted. The acceptance of the model establishes a related degree of measurement invariance for the original model (Milfont & Fischer, 2010).

In order to test the three-factor model for measurement invariance, this study used the first two levels of invariance testing: the configural invariance and the metric invariance (Milfont & Fischer, 2010). The configural invariance tests if the model's configuration/structure is consistent across all groups. In other words, does the factor structure hold across these groups? In order to test this invariance, the model is restricted by constraining its factor structure. The metric invariance compares different groups' responses to the items in order to test if they are identical or not. Namely, it tests whether items' factor loadings are the same across all groups. Therefore, to test for this invariance, the model is restricted through the constraining of all factor loadings to be the same in all groups.

The categorical variable used for defining groups was the participants' gender. The configural and metric invariances were tested by creating two groups: females and males. The results of these invariance tests are listed in Table 2.15.

Table 2.15
Fit Indexes for Invariance Tests for the Three-Factor Model

Model	χ^2	df	χ^2 (diff)	df (diff)	RMSEA	SRMR	NFI	CFI	TLI	Passed
Configural	626.66	102	–	–	.060	.053	.92	.93	.91	Yes

Metric	636.92	114	10.25	12	.057	.053	.92	.93	.92	Yes

The configural invariance test for the three-factor model produced acceptable fit index values (RMSEA = .06, SRMR = .053, NFI = .92, CFI = .93, TLI = .91). The fit indexes generated by the metric invariance test were also acceptable: RMSEA = .057, SRMR = .053, NFI = .92, CFI = .93, TLI = .92. Lastly, the chi-squared difference test comparing the metric model to the configural model was not significant (χ^2 = 10.25, df = 12, p = .59), which indicated no significant difference between two models. These results demonstrated that the three-factor model satisfied the requirements of measurement invariance across gender.

Instrument's reliability and validity. Tavakol & Dennick (2011) asserted that two important elements, validity and reliability, are crucial for evaluating survey instruments. Furthermore, they argued that the instrument cannot be valid unless it is reliable and reliability estimates are also used for measuring an error in a test. Also, researchers stated the importance of making sure the instrument passes content and construct validities (Moskal, Leydens & Pavelich, 2002).Therefore, ensuring the high degree of validity was the focus of the SIEVEA instrument.

The validity indicates the extent to which the survey measures whatever construct it intends to measure (Groves, Fowler, Couper, Lepkowski, Singer, & Tourangeau (2009). Shadish, Cook and Cambell (2002, p.20) made distinctions between the two types of causal generalizations: "[…] construct validity generalizations (inferences about the constructs that research operations represent) and external validity generalizations (inferences about whether the causal relationship holds over variation in persons, settings, treatment, and measurement variables)."

Since the SIEVEA instrument was intended for measuring multiple constructs, it was important to verify the construct validity by measuring whether convergent and discriminant validities were satisfied (Agarwal, 2011). According to Trochim (2006), it is not adequate to attest either convergent or discriminant validities of the instrument in order to ensure its construct validity is satisfied; both convergent and discriminant validities should be confirmed. For that purpose, correlations between constructs' items were measured (Trochim, 2006).

Jöreskog (1967) highlighted the importance of assessing the instrument's construct validity during a CFA. Measuring and confirming the instrument's validity is a crucial step in analyzing the instrument's psychometric properties and must be carried out before the instrument's data can be used for other statistical tests. Two major types of validity, convergent and discriminant, need to be estimated for supporting the evidence of construct validity (Campbell & Fiske, 1959). According to these scholars, convergent validity indicates whether the constructs are well measured by their respective items or not. Likewise, discriminant validity assesses the degree to which measures of different constructs are unrelated.

Fornell-Larcker (1981) proposed several convenient and widely used measures for establishing validity and reliability. The Composite Reliability (CR) and the Average Variance Extracted (AVE) are used for assessing reliability and convergent validity. CR is similar to Cronbach's alpha, but is considered less biased, as Cronbach's alpha tends to underestimate true reliability (Peterson & Kim, 2013). CR has an acceptable value of .7 and above. The AVE is similar to explained variance in EFA. It measures the average variance in items that a construct is managed to explain. In other words, the AVE expresses the level of variance captured by a construct versus the level due to measurement errors. AVE values .5 and above are considered acceptable for establishing convergent validity.

The discriminant validity can be evaluated using the Maximum Shared Variance (MSV) and the Average Shared Variance (ASV), which respectively measure maximum and average variances among constructs. Both the MSV and the ASV are expected to be lower than the AVE for all constructs for confirming discriminant validity (Hair, Black, Babin, & Anderson, 2010). Furthermore, the square root of the AVE should be greater than inter-construct correlations for all constructs.

The CR, AVE, MSV, and ASV were calculated for the three-factor model in order to assess its convergent and discriminant validities. AMOS software (Arbuckle, 2016) was used to produce input data (factor correlations and standardized regression weights) for calculating these metrics and assessing the convergent and discriminant validities. Table 2.16 contains the results of these calculations.

Table 2.16
Validation Measures and Inter-Construct Correlations (the Three-Factor Model)

Construct	CR	AVE	MSV	ASV	C1	C2	C3
C1	.882	.521	.393	.311	**.722**	–	–
C2	.755	.510	.260	.245	.479	**.714**	–
C3	.727	.574	.393	.327	.627	.510	**.758**

Note. The bolded numbers on the diagonal on the right part of the table are the square roots of the AVE. The non-diagonal numbers are inter-construct correlations.

As the data in Table 2.16 shows, the three-factor model satisfied all reliability and validity requirements. Indeed, since the CR values of all three constructs were greater than the threshold value of .7, the reliability requirement was met. Likewise, convergent validity was confirmed as well, because all AVE values exceeded .5. Lastly, all discriminant validity requirements were satisfied: both the MSV and the ASV were less than the AVE for all three constructs, and the square roots of the AVE (the bolded numbers on the diagonal on the right part of the table) were greater than the corresponding inter-construct correlations.

Figure 2.12 shows all major steps of CFA of SIEVEA.

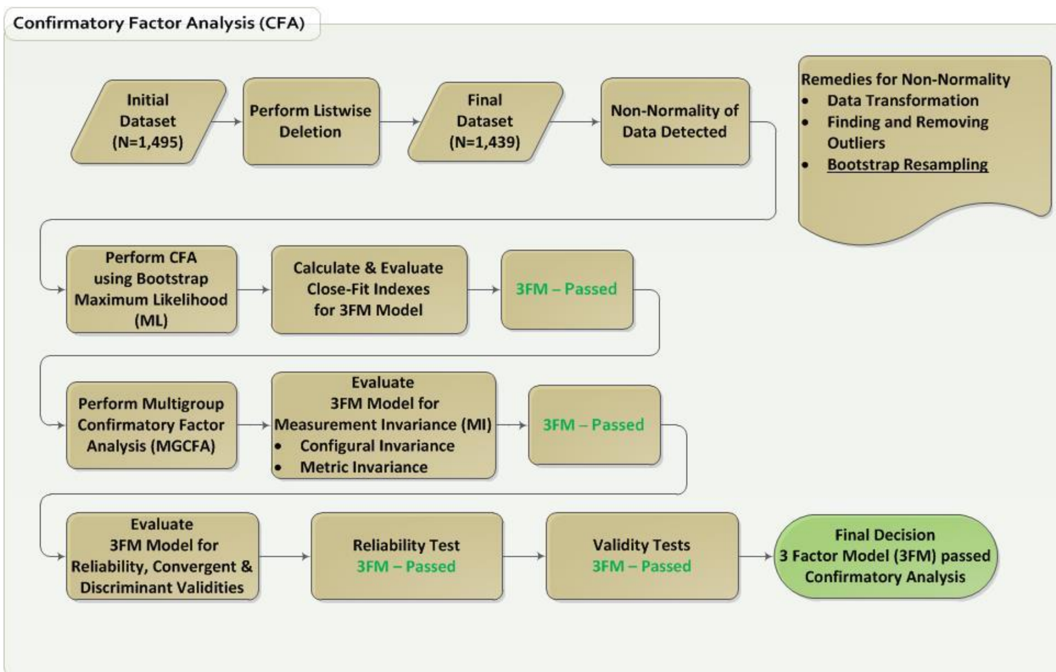


Figure 2.12. Confirmatory factor analysis of SIEVEA

Discussion. The EFA conducted during the previous study produced three candidate models as representations of the factor structure of data collected using the SIEVEA survey instrument. Further analyses of these models evaluated their fit to data, their alignment with the research constructs, and their factor loadings. Because of these analyses, the three-factor model was selected as the final model.

The PCFA analysis carried out in this study indicated that the three-factor model was a good fit. Therefore, it was eligible for CFA. In the second part of this study, this model was analyzed by CFA using a new data set exclusively collected to conduct confirmatory factor analysis. The three-factor model showed an acceptable fit and, therefore, its choice as the correct model was confirmed.

The factor loadings produced by CFA (see data in Tables 2.13) were greater than those produced by EFA for almost all items. This result confirmed once more the validity of factor structures suggested by EFA, and provided a proof of the survey's configural validity. Moreover, the results of CFA were even more remarkable, as the confirmatory analyses used a completely new sample collected independently from the original data. Using a new data set provided extra credence to the confirmatory process and strengthened the argument of survey's validity.

The three-factor model is appropriate for measuring the following latent constructs: students' science identities and motivation (C1), environmental attitudes (C2), and science values (C3). The model has a good fit and a simple factor structure. The three-factor model performed well during measurement invariance testing while using gender as a grouping variable. This result shows that data collected using the SIEVEA survey can be confidently implemented in research and statistical tests that will compare and contrast construct scores between females and males. The three-factor model can be used for this type of research and tests.

The three-factor model was also tested for reliability. Additionally, this model was assessed for construct validity. These tests indicated that the three-factor model was reliable and passed all the requirements of convergent and discriminant validities. Table 2.17 provides a summary of various test and validation results for the three-factor model.

Table 2.17
Summary of Fit Tests and Validations

Test Type	Three-Factor Model
Model Fit	Passed
Measurement Invariance	Passed
Reliability	Passed
Convergent Validity	Passed
Discriminant Validity	Passed

As a result of this study, the three-factor model was fully validated and was confirmed as an apprioriate representation of the SIEVEA's factor structure. Indeed, the three-factor model had an acceptable fit to data, high levels of reliability and construct validity. Therefore, all future studies using the data collected by the SIEVEA instrument will utilize the three-factor model.

Limitations and recommendations for future research. It should be noted that the science values factor has only two items in the three-factor model. Because of this, this factor's usefulness in capturing its latent variable is somewhat limited.

Most psychometric analyses like the Rasch analysis do not work well for constructs with less than three items. Therefore, while performing parametric statistical tests on the science values construct using collected data, it is important to note this limitation, and a certain level of skepticism should accompany the test results.

It is recommended to improve the SIEVEA survey instrument by adding more items related to the science value construct. This will overcome the survey's current limitation by enhancing its reliability in capturing the science value latent variable. Since the survey was not big and Item #9 was removed during the EFA, adding a few more items should not be a problem: the survey will continue to be fairly short and convenient to complete.

Item wording is another area of possible revisions regarding instrument and item design. The SIEVEA included only positively worded items, which is the most commonly used type of items in many instruments (Usher & Pajares, 2009). Some researchers recommended mixing positively and negatively worded items in the same instrument in order to improve the instrument's ability to distinguish between extreme and moderate respondents on the construct and lessen the possibility of ceiling and basement effects (Spector, Van Katwyk, Brannick, & Chen, 1997).

However, the usefulness of negatively worded items is not universally accepted. Marsh (1996) claimed that item wording reflects more on artifacts rather than on constructs. Other scholars argued caution when using negatively worded items as, depending on the type of negations (e.g., polar opposite vs. item reversal), the instrument's reliability and validity can stay intact or get compromised (Schriesheim, Eisenbach & Hill, 1991). Indeed, responses on the negatively worded items might solely be an indication of personal bias or style (DiStefano & Motl, 2006).

Still, it is worth trying to revise some items in order to have several negatively worded items along with positively worded ones. Since some researchers claim that positively and negatively worded items load differently (Ray, Frick, Thornton, Steinberg, & Cauffman, 2016), it will be interesting to determine whether the inclusion of negatively worded items will produce the same factor structure or impact SIEVEA's reliability and validity.

Another recommendation is to expand the geography of data collection. Even though the survey was conducted in multiple states (New Jersey, Pennsylvania and Connecticut), its online nature made it possible to collect data in various states, as well as outside of the U.S.A. This allows for more diverse data collection, leading to improved external validity of future studies.

References

Agarwal, N. K. (2011). Verifying survey items for construct validity: A two stage sorting procedure for questionnaire design in information behavior research. *Proceedings of the American Society for Information Science and Technology, 48*(1), 1-8.

Arbuckle, J. L. (2016). *Amos user's guide version 24*. Chicago, IL: SmallWaters.

Bandalos, D. L. (1996). Confirmatory factor analysis. In J. Stevens (Ed.), *Applied multivariate statistics for the behavioral sciences* (3rd ed., pp. 389-420). Mahwah, NJ: LEA.

Baumgartner, H., & Hombur, C. (1996). Applications of structural equation modeling in marketing and consumer research: A review. *International Journal of Research in Marketing*, *13*, 139-161.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107,* 238-246.

Bentler, P. M. (1995). *EQS structural equations program manual.* Encino, CA: Multivariate Software.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88,* 588-606.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York, NY: Wiley.

Bollen, K. A. (1989). A new incremental fit index for general structural equation models. *Sociological Methods & Research*, *17*(3), 303-316.

Bollen, K. A., & Stine, R. A. (1992). Bootstrapping goodness-of-fit measures in structural equation models. *Sociological Methods & Research, 21,* 205-229.

Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. New York, NY: Guilford Press.

Browne, M. W. (1990). *MUTMUM PC user's guide*. Columbus: Ohio State University, Department of Psychology.

Browne, M. W., & Cudek, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing Structural Equation Models* (pp. 136–162). Newbury Park, CA: Sage.

Byrne, B. M. (2013). *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. New York, NY: Routledge.

Campbell, D. T. & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56-81.

DiStefano, C., & Motl, R. W. (2006). Further investigating method effects associated with negatively worded items on self-report surveys. *Structural Equation Modeling*, *13*(3), 440-464.

Eccles, J. S., & Wigfield, A. (1995). In the mind of the actor: The structure of adolescents' achievement task values and expectancy-related beliefs. *Personality and Social Psychology Bulletin*, *21*(3), 215-225.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics, 7*, 1–26.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*(3), 272-299.

Fornell, C. & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 39-50.

Gignac, G. E. (2009). Partial confirmatory factor analysis: Described and illustrated on the NEO–PI–R. *Journal of Personality Assessment*, *91*(1), 40-47.

Groves, R. M., Fowler Jr, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey methodology* (Vol. 561). Hoboken, NJ: John Wiley & Sons.

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2010). *Multivariate Data Analysis* (7th Eds.). New York, NY: Pearson.

Henson, R. K., & Roberts, J. K. (2006). Use of exploratory factor analysis in published research common errors and some comment on improved practice. *Educational and Psychological Measurement*, *66*(3), 393-416.

Hortensius, L. (2012). *Project for introduction to multivariate statistics: Measurement invariance*. Retrieved from http://www.tc.umn.edu/~horte005/docs/MeasurementInvariance.pdf

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1-55.

Ichikawa, M., & Konishi, S. (1995). Application of the bootstrap methods in factor analysis. *Psychometrika, 60*, 77–93.

Jackson, D. L., Gillaspy Jr, J. A., & Purc-Stephenson, R. (2009). Reporting practices in confirmatory factor analysis: An overview and some recommendations. *Psychological methods*, *14*(1), 6-23.

Jöreskog, K. G. (1967). A general approach to confirmatory maximum likelihood factor analysis. *ETS Research Bulletin Series*, 183-202.

Kaiser, H.F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, *20*, 141-151.

Kline, R. B. (2005). *Principles and Practice of Structural Equation Modeling*, (2nd ed.). New York: Guilford.

Lent, R. W., Lopez, F. G., Brown, S. D., & Gore Jr, P. A. (1996). Latent structure of the sources of mathematics self-efficacy. *Journal of Vocational Behavior*, *49*(3), 292-308.

Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling*, *9*(2), 151-173.

Little, T. D., Rhemtulla, M., Gibson, K., & Schoemann, A. M. (2013). Why the items versus parcels controversy needn't be one. *Psychological Methods*, *18*(3), 285.

MacCallum, R. C., & Austin, J. T. (2000). Applications of structural equation modeling in psychological research. *Annual Review of Psychology*, *51*(1), 201-226.

Marsh, H. W. (1996). Positive and negative global self-esteem: A substantively meaningful distinction or artifactors? *Journal of Personality and Social Psychology*, *70*(4), 810.

Matsunaga, M. (2010). How to factor-analyze your data right: Do's, don'ts, and how-to's. *International Journal of Psychological Research*, *3*(1), 97-110.

Milfont, T. L., & Fischer, R. (2010). Testing measurement invariance across groups: Applications in cross-cultural research. *International Journal of Psychological Research*, *3*(1), 111-130.

Moskal, B. M., Leydens, J. A., & Pavelich, M. J. (2002). Validity, reliability and the assessment of engineering education. *Journal of Engineering Education*, *91*(3), 351-354.

Nesselroade, J. R. (1994). Exploratory factor analysis with latent variables and the study of processes of development and change. In A. von Eye & C. C. Clogg (Eds.), *Analysis of latent variables in developmental research* (pp. 131-154). Newbury Park, CA: Sage.

Nevitt, J., & Hancock, G. R. (2001). Performance of bootstrapping approaches to model test statistics and parameter standard error estimation in structural equation modeling. *Structural Equation Modeling*, *8*(3), 353-377.

Peterson, R. A., & Kim, Y. (2013). On the relationship between coefficient alpha and composite reliability. *Journal of Applied Psychology, 98*(1), 194.

Pett, M. A., Lackey, N. R., & Sullivan, J. J. (2003). *Making sense of factor analysis: The use of factor analysis for instrument development in health care research*. Thousand Oaks, CA: Sage.

Ray, J. V., Frick, P. J., Thornton, L. C., Steinberg, L., & Cauffman, E. (2016). Positive and negative item wording and its influence on the assessment of callous-unemotional traits. *Psychological Assessment*, *28*(4), 394.

Reise, S. P., Waller, N. G., & Comrey, A. L. (2000). Factor analysis and scale revision. *Psychological Assessment*, *12*(3), 287.

Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. *Psychological Bulletin*, *114*(3), 552.

Russell, D. W. (2002). In search of underlying dimensions: The use (and abuse) of factor analysis in Personality and Social Psychology Bulletin. *Personality and Social Psychology Bulletin*, *28*(12), 1629-1646.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*(2), 147-177.

Schriesheim, C. A., Eisenbach, R. J., & Hill, K. D. (1991). The effect of negation and polar opposite item reversals on questionnaire reliability and validity: An experimental investigation. *Educational and Psychological Measurement*, *51*(1), 67-78.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Belmont, CA: Wadsworth Cengage Learning.

Spector, P. E., Van Katwyk, P. T., Brannick, M. T., & Chen, P. Y. (1997). When two factors don’t reflect two constructs: How item characteristics can produce artifactual factors. *Journal of Management*, *23*(5), 659-677.

Stevens, J. (1996). *Applied multivariate statistics for the social sciences* (3rd ed.). Mawah, NJ: Lawrence Erlbautn Associates.

Stevens, T., Olivárez, A., & Hamman, D. (2006). The role of cognition, motivation, and emotion in explaining the mathematics achievement gap between Hispanic and White students. *Hispanic Journal of Behavioral Sciences*, *28*(2), 161-186.

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, *2*, 53-55.

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

Trochim, W. M. (2006). *The Research Methods Knowledge Base*, 2nd Edition. Retrieved from http://www.socialresearchmethods.net/kb/

Tsagkanos, A. G. (2007). A bootstrap-based minimum bias maximum simulated likelihood estimator of mixed logit. *Economics Letters*, *96*(2), 282-286.

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38*, 1-10.

Usher, E. L., & Pajares, F. (2009). Sources of self-efficacy in mathematics: A validation study. *Contemporary Educational Psychology, 34*(1), 89-101.

Van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology, 9*(4), 486-492.

Williams, L. J., & O'Boyle, E. H. (2008). Measurement models for linking latent variables and indicators: A review of human resource management research using parcels. *Human Resource Management Review*, *18*(4), 233-242.

Study 1C – Validation of SIEVEA using the Rasch Analysis

Introduction. Exploratory and confirmatory factor analyses are prevalent statistical methods for respectively discovering and confirming the factor structure present in survey data. These factors reflect the underlying latent variables that the survey was created to measure. One major difficulty with measuring latent variables using surveys is the rating (ordinal) scale nature of the survey data. Indeed, since this data is not interval scale, the direct application of common parametric statistical tests like the t-test or ANOVA to survey data is not appropriate (McCrum-Gardner, 2008). Regrettably, many researchers neglected to consider the ordinal nature of survey data when deciding how to conduct parametric statistical tests; they simply used raw survey scores for these tests (Boone, Townsend & Staver, 2010). In order to conduct proper parametric statistical tests on survey data, it is necessary to convert the data from ordinal to interval scale (Boone & Scantlebury, 2006).

Additionally, the factor scores generated by factor analysis are based on the classical test theory also known as the true score theory. According to this theory, the respondent’s observed score on a survey item is the sum of his/her true score (person’s ability score) and an error (luck or random error) (Lord, 1980). While this theory provides a simple and intuitive approach to survey and test scoring, it has several serious limitations. First, in this theory, it is not possible to separate respondent’s characteristics, like the person’s individual’s ability from test characteristics like item difficulty. Next, the classical test theory defines reliability as "the correlation between test scores on parallel forms of a test" (Hambleton, Swaminathan, & Rogers, 1991). However, there is no consensus among researchers regarding what parallel tests are and whether there is a practical way to carry out these tests. Another shortcoming of this theory is its assumption that the standard error of measurement is the same for all respondents. Lastly, the classical test theory is test-oriented, not item-oriented. As a result, it cannot predict how a specific participant will perform on a survey or test item (Hambleton, Swaminathan, & Rogers, 1991).

Due to the above-mentioned shortcomings of the classical test theory, more sophisticated theories were developed, like the Item Response Theory (IRT), in psychometrics. The Rasch model, created by Danish mathematician Georg Rasch (1960), is a probabilistic model belonging to the IRT family of models. It is the method of choice of many researchers for designing, validating, using and improving survey instruments. The Rasch model is widely used in math and science research education (Sjaastad, 2014; Zain, Samsudin & Jusoh, 2010), in medical science (Gothwal, Wright, Lamoureux & Pesudovs, 2009), in health science (Wang et al., 2006) and in various medical disciplines (American Dental Association, American Board of Pediatric Dentistry, Medical Examiners, American Society of Clinical Psychologists, American Board of, Boone & Scantlebury, 2006) for designing the survey instruments and measuring and enhancing their psychometric quality. Moreover, the Rasch analysis was employed for evaluating large datasets in the *Program for International Student Assessment* (PISA, Bond & Fox, 2007), the *Trends in International Mathematics and Science Study* (TIMSS, Sjaastad, 2014), the *International Civic and Citizenship Education Study* (ICCES, Schulz & Fraillon, 2011), the *Lexile framework* (Stenner, Burdick, Sanford, & Burdick, 2006) and the *National Assessment of Educational Progress* (NAEP, Lee, 2004).

This study conducted the Rasch analysis on the SIEVEA survey instrument in order to archive the following goals: validation of the SIEVEA instrument, conversion of the ordinal scores of SIEVEA's data to interval scale to prepare for conducting of parametric statistical tests, and exploration of SIEVEA's psychometric properties to generate ideas regarding how this instrument can be enhanced.

Overview of the Rasch analysis. The Rasch model is a single-parameter model, whereas most other models of the IRT family are two- or three-parameter models (Boone & Scantlebury, 2006). Like exploratory and confirmatory factor analysis, the Rasch model is used for measuring latent variables. However, unlike factor analysis, the Rasch analysis takes under the consideration "person ability" and "item difficulty" strands. While those concepts are self-explanatory when applied to knowledge-based tests, they have a different meaning for attitude and opinion measuring instruments. Sjaastad (2014, p. 214) explained that "… ability refers to the amount of a property possessed by a person" and "… difficulty refers to the amount of a latent variable a person must possess to have a 50% chance of receiving score on the item." During the Rasch analysis, both item difficulties and person abilities are placed on a scale called logit with higher scores indicating higher abilities or difficulty and vice versa (Gothwal et al., 2009).

The Rasch model has two different versions: dichotomous model and polytomous model (Sjaastad, 2014). The dichotomous model is used for analyzing instruments with two possible response categories (true/false, yes/no, agree/disagree, etc.), whereas the polytomous model is applicable to instruments with more than two response categories (5-point Likert scale, 7-point Likert scale, etc.). The mathematical form of the dichotomous Rasch model is given by the following equation:

$$P(\text{Score} = 1) = \frac{e^{(B-D)}}{1 + e^{(B-D)}}$$

Here P is the probability of a person with ability B to answer correctly (score = 1) an item with difficulty D. The polytomous model uses a similar equation which gives the probability of receiving a specific score given a person's ability and an item's and its steps' difficulties (Sjaastad, 2014).

The Rasch analysis provides researchers with opportunities to analyze ordinal scale data (Mallinson, 2007), which would otherwise be impossible with other types of analyses. Researchers who encouraged the use of the Rasch analysis argued that traditional analyses incorrectly treated the item responses as interval data (Wang, Yao, Tsai, Wang, & Hsieh, 2006). Contrarily, the Rasch technique allows for the conversion of item responses from ordinal scales to interval measures (Wright & Stone, 1979). Although researchers may spend more time performing the Rasch analysis than carrying out traditional analysis, the outcome is promising as it provides deeper understanding, accurate results, and information about the strengths and weaknesses of the implemented instrument (Boone & Scantlebury, 2006). Additionally, it is important to transform ordinal data from the Likert scale survey into interval scale data prior to conducting parametric statistical tests like the t-test and ANOVA.

In contrast to the CFA, which focuses on the covariance between test items, the item response theory (IRT), including the Rasch model, aims to examine item responses (Reise, Widaman & Pugh, 1993). Similar to the classical test theory, the IRT considers the observed score of the measured variable as the true score and the difference between true score and observed score as the error of measurement (Lord, 1980). However, as Lord argued, there is neither a bias nor correlation between the error of measurement and the true score in the IRT model, whereas in the classical test theory, raw scores are always biased (Wright, 1997). Although the classical test theory very often produces results similar to the Rasch analysis, the Rasch theory still possesses characteristics that the classical test theory lacks, such as handling the missing data, computation of an individual's fit statistics, locating persons' ability and the items' difficulty on the same scale of measurement (Wright, 1992), transforming of ordinal level data to interval level data (Rasch, 1980) and so on.

Additionally, the Rasch analysis provides instrument developers and research practitioners with different methods for measuring the instrument's reliability and validity (Boone, Townsend & Staver, 2010). It can be used to validate the instrument by checking how well all the items work together in representing their underlying construct and, therefore, provide additional construct validation (Fortus & Vedder-Weiss, 2014). Moreover, the Rasch analysis is very powerful tool in the following areas: determining how well the instrument works for a specific pool of respondents; assessing whether the instrument properly separates the respondents based on their ability; finding redundant response categories; checking whether the instrument is bias free; identifying which responders the instrument does not provide reliable measure for; finding the instrument's problematic areas and identifying opportunities for its enhancement (Boone et al., 2010; Sjaastad, 2014). In addition, often the Rasch analysis is used not only for measuring the instrument's reliability and validity, but also for reducing the number of test items by removing the redundant items. This makes the instrument shorter yet maintains the appropriate accuracy (Hibbard, Mahoney, & Stockard, 2005).

Performing the Rasch analysis usually involves scrutinizing the following modeling properties of the instrument: analyses of test targeting, person separation, item fit, person fit, differential item functioning, functioning of response categories and tests of unidimensionality (Sjaastad, 2014). The following discussions briefly examine all these areas.

Test targeting. The instruments can be either well-targeted or poor-targeted. In a well-targeted instrument, item difficulties are expected to be approximately even-spaced and well fitted on the scale like the marks on the ruler with persons' abilities located on the opposite side of the ruler (Mallinson, Stelmack, & Velozo, 2004). Moreover, in a well-targeted instrument, most item difficulties should be located in the same region of the scale as the abilities of the persons taking the instrument (Sjaastad, 2014). Indeed, if item difficulties are mostly located outside of the respondents' ability range (either above or below it), then respondents will either provide mostly correct answers or incorrect answers on these items. This will make them essentially redundant and damage the instrument's ability to properly differentiate respondents based on their ability (Weller, Dieckmann, Tusler, Mertz, Burns, & Peters, 2013).

Furthermore, Sjaastad argued that the instrument should include a few items located in the lower and higher regions of the scale in order to help in identifying persons with low and high abilities. Therefore, a good instrument should include items at various difficulty levels (Gothwal et al., 2009). Also, the reliable instrument should not have significant ceiling or floor effects (Cappelleri, Lundy, & Hays, 2014), which can happen if there are a significant number of respondents whose scores are either greater than the maximum or less than the minimum score the instrument can measure.

Person separation. The item and person separation statistics are valuable analytical tools in the Rasch analysis, as they can be used to evaluate the instrument's reliability and monitor its continuing effectiveness. Person separation shows how effective the instrument is in separating persons who are being measured.

Likewise, item separation shows how well the sample of people can separate items used in the instrument. The statistics provided by the Rasch analysis indicate how well the test separates respondents and ranks them (Sjaastad, 2014), thus allowing assessment of the reliability of the instrument. One of these reliability statistics is the person separation index (PSI) (Cappelleri et al., 2014). The PSI is analogous to Cronbach's alpha and their values are often close when applied to the same sample. However, there is an important distinction between these two metrics: the PSI uses logit values, whereas Cronbach's alpha is calculated on raw scores (Tennant & Conaghan, 2007). Since the items with difficulty estimates close to the person's ability provide the most information about the person, it can be inferred that the person separation index depends on the quality of the test targeting (Sjaastad, 2014). The PSI values range from 0 to 1 with values close to 1 indicating a good separation.

Person separation can also be assessed by inspecting the Wright map of the construct (Sjaastad, 2014). The Wright map is an aggregate map of all respondents' proficiency levels versus all the item difficulties, placed on the same logit scale. Typically, the Wright map has 3 columns. The first column is simply the logit scale ranging from -6 to 6 or -4 to 4. The second column depicts the histogram of respondents' abilities/proficiencies. Lastly, the third column shows the Thurstonian thresholds of all item steps, where a Thurstonian threshold is the location in logits at which a respondent has a 50% chance of achieving a score in that step's category or higher (Wilson, 2005). The Wright map is also a handy tool for evaluating the instrument's test targeting and discovering any ceiling or floor issues.

Unidimensionality assumption of the Rasch model. The Rasch model assumes that the analyzed data is unidimensional, namely, that it measures a single underlying trait (Gothwal, et al., 2009). However, when conducting the Rasch analysis on survey data, either the multidimensional approach can be used, meaning the data is analyzed as a whole (Wang et al., 2006), or each construct present in a questionnaire can be measured separately as a single latent trait (Andrich, 1988; Brentari & Golia, 2007) while performing "… one test at a time" (Wang et al., 2006, p. 608).

Unidimensionality or multidimensionality of data can be established by either conducting factor analysis or Rasch analysis. Some researchers suggested using a multidimensional approach in order to take correlations between the latent traits into account and make the measurement more precise (Wang et al., 2006). Furthermore, Wang and colleagues claimed that the multidimensional approach works better with short tests, because it allows using the whole data for all dimensions.

Since the SIEVEA had only 15 items, it made sense to consider using the multidimensional approach in addition to the unidimensional one. However, since the data was already factor analyzed during prior studies and as all latent variables were identified, the simpler unidimensional approach for performing Rasch analysis on the subsets of data representing separate latent variables was implemented.

Figure 2.13 lists the key concepts and research methods used in evaluating instrument's test targeting, person separation and unidimensionality.

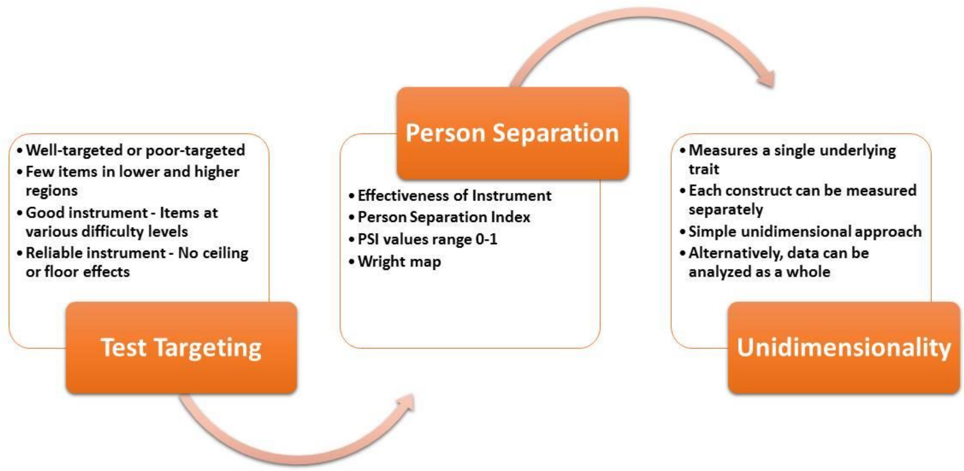


Figure 2.13. Test targeting, person separation and unidimensionality

Item fit. Item fit statistics, as the word suggests, indicates whether the individual items fit within the remaining items or not. Furthermore, the person fit statistics have an ability to predict students' test answers while evaluating students' abilities (Boone & Scantlebury, 2006). Because of the Rasch model's unidimensionality assumption, it makes no measurement sense to include misfit items in the model, leading to their exclusion from the general pool of item analysis (Boone & Scantlebury, 2006).

Calculating the item's fit is quite simple. First, all respondents are divided into groups based on their abilities. Second, these groups' means to the item are calculated. Lastly, the difference, called the fit residual, between the observed group means and the expected values is calculated and used as an indicator of the item fit (Sjaastad, 2014). Depending on the value of the fit residual, the item may have *good fit*, *under-discrimination* or *over-discrimination*. If the observed values are close to the expected values, then the item has a good fit. If the observed values do not increase as much as the ability increase would suggest, then the item under-discriminates. Likewise, over-discrimination happens when the item separates persons in a small region of the logit scale but provides little information about persons whose abilities are outside of that region.

The item's fit can also be visually inspected by plotting the mean scores on the item characteristics curve (ICC). The ICC displays the probability of a person's scoring on the item (for a dichtomous response) or responding to a particular category of an item (for a polytomous response) based on the individual's ability. In the case of a polytomous response, the ICC is sometimes called a category response or a probability curve (Cappelleri et al., 2014). The category probability curves allow researchers to interpret the rating scale model's responses. The x-axis of the curve is the independent variable indicating a person's ability on a particular subscale (e.g., science identity, expectation of success in science, values of science, environmental attitudes), whereas the y-axis is a dependent variable that represents the probability of achieving a particular score/category on the item (e.g., categories 0-4 in case of 5-point Likert scale). Due to probabilistic assumptions of the Rasch model, the ICC graph and category probability curves are not linear, but rather logistic regression lines (Sjaastad, 2014).

For each individual item, the goodness of fit assessment could be calculated by utilizing the Mean Square error statistics (*MNSQ)*, which is the average of the squared residuals (Wright & Masters, 1982). The two main types of mean squares are *infit MNSQ* and *outfit MNSQ* that differ from each other in the terms of weights assigned to person scores (Ding, 2011). When calculating the infit statistics, more weights are assigned to those persons' scores whose ability levels are close to the item difficulty. On the other hand, all person scores have the equal weight for calculating the outfit mean square statistics. Because of this, the outfit MNSQ and the infit MNSQ have different sensitivities regarding the respondents' behavior. Outfit MNSQ is sensitive to respondents' unexpected behavior on items that are far from the person's proficiency level, whereas infit MNSQ is sensitivity to unexpected behavior on items close to the person's proficiency level (Wang et al., 2006). Therefore, the outfit statistics is more sensitive to outliers than the infit is.

According to Wang et al. (2006), the infit and outfit MNSQ statistics have an expected value of 1, provided the data fits the model. Values less than 1 indicate too many predictable observations and redundancy; this is also referred to as over-fit. Similarly, lots of unpredictable observations cause these values to be greater than 1 and are classified as under-fit (also called misfit (Eckes, 2005)). Generally, if an item's infit and/or outfit mean squares fall between .7 and 1.3, the item is considered a good fit under a unidimensional construct (Bond & Fox, 2007). When analyzing survey's data, MNSQ values between .6 and 1.4 indicate a reasonable fit (Wright, Linacre, Gustafson, Martin-Lof, 1994). Moreover, Linacre (2012) suggested using a less stringent range of .5 to 1.5 to indicate a good fit. However, since the use of these boundary values is a heuristic rather than a rule, some researchers recommend using ICC for evaluating the item fit, in addition to the infit and outfit statistics (Schulz & Fraillon, 2009).

Person fit. The person fit indicates whether survey or test respondents' answers fit the model or not. This fit measure is based on the assumption that the individuals are more likely to provide correct responses on easy items than on difficult items (Sjaastad, 2014). Furthermore, the higher ability of the respondent in comparison to the difficulty of the item increases chances of a correct response (Weller et al., 2013).

Therefore, it is expected that the individual will correctly answer most items with difficulties less than his/her ability and will fail on most items with difficulties exceeding his/her ability. For items with difficulties close to the respondent's ability, the response pattern will be more random: some correct answers, some incorrect answers. The Rasch analysis evaluates each individual's responses based on its closeness to the expected response pattern (Boone & Scantlebury, 2006). If responses are very close to the expected pattern, that person over-fits the model. Inversely, if responses are too random (lots of misses on easy items and lots of correct answers on difficult items), then the person under-fits the model.

The person fit statistic provides a numeric value that can be used to decide whether the person's responses fit the model or not. Values close to 0 indicate good fit. On the other hand, values below -2.5 show evidence of over-fit, which can happen if the test consists of mostly easy or mostly difficult questions for the person. Likewise, values above +2.5 are indicators of under-fit; the person gave many incorrect answers to easy questions but was able to correctly answer lots of difficult questions (Sjaastad, 2014).

The infit and outfit MNSQ statistics can also be used to measure the fitness of the person's response. Values between .5 and 1.5 indicate a good fit (Linacre, 2012). Bond and Fox (2001) suggested stricter values: .7 for lower bound and 1.3 for upper bound.

Interestingly, using the person fit statistics quantitatively reveals useful information about respondents' answers, which would otherwise have required qualitative analysis (Boone & Scantlebury, 2006). For example, the person fit statistics can help in uncovering various response patterns hidden in the data set, like identifying common problems with cheating and guessing (Sjaastad, 2014) or figuring out whether a particular group of respondents miss items (Boone & Scantlebury, 2006).

Differential item functioning. Differential item functioning (DIF) happens when persons with the same amount of the latent trait respond differently to a particular item (Gothwal et al., 2009). For example, DIF with respect to gender means that the item favors respondents of one gender (e.g., females) over respondents of the other gender (e.g., males). It can be said that this type of item is biased toward one group of respondents and biased against another group.

Boone and Scantlebury (2006) considered the identification of biased items to be a very important component of instrument evaluation. DIF will indicate whether certain items were much easier to answer by males or vice versa. As Boone and Scantlebury stated, DIF serves as face validity for identifying a test bias. Also, the program looks for the test items within a subgroup and compares its location to the other test items'. If discrepancies are found to be statistically significant, then it is a good indication of a biased test item. Very often, items possessing DIF characteristics are removed from the test before the calculation of students' mean scores.

Functioning of response categories. Research claims that in order to achieve high reliability, validity and discriminating power, the rating scales used by the survey should have at least five response categories (Preston & Colman, 2000). Even though 5-point Likert-type psychometric scale satisfies this requirement and is widely used by researchers, its popularity of is not fully justified.

Criticism is mostly directed toward the mid-category (category 3; “Neither Agree Nor Disagree”). It turned out that the middle response “… is at least sometimes used as a ‘dumping ground’ for unsure or non-applicable responses” attracts respondents who do not care about the answers (Kulas, Stachowski, & Haynes, 2008, p. 258). Some scholars believe that respondents choose the neutral response because of their uncertainty of answering that specific item (DeMars & Erwin, 2004). Additionally, some response categories may be redundant. This can happen when the increase of ability causes a jump in category by taking the respondent from one category (e.g., category 2) to the higher category (e.g., category 4) and skipping the category in-between (e.g., category 3) (Sjaastad, 2014).

The Rasch analysis allows researchers to assess whether the response categories were appropriate or not. This can be accomplished by inspecting the response scales’ probability curves (ICCs). On these probability curves, the thresholds, indicating advancement from one category to the next category, are located at the intersections between the curves of two adjacent response categories. If the thresholds follow the same order as the categories, then the response category is appropriate without redundancies. Otherwise, the “reversed” or “disordered” thresholds indicate problems with the appropriateness of their response categories (Sjaastad, 2014).

Figure 2.14 lists the key concepts and research methods used in evaluating instrument's item fit, person fit, DIF and functioning of response categories.

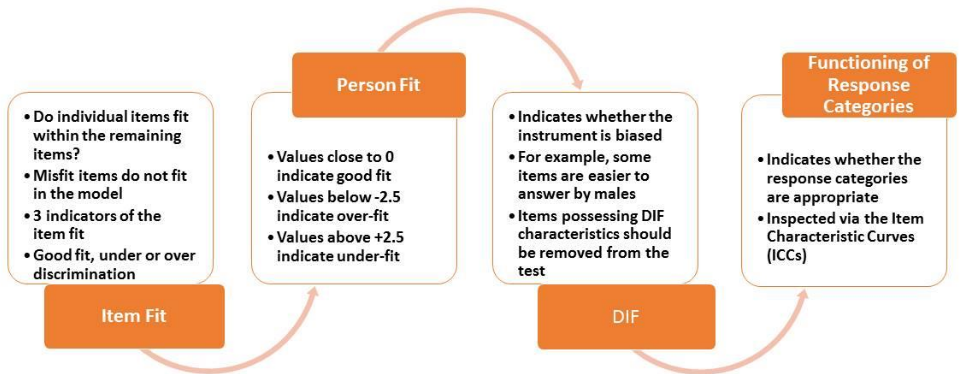


Figure 2.14. Item fit, person fit, DIF and functioning of response categories

Method. ***Instrument.*** The survey instrument SIEVEA was developed for measuring high school students' science identities, expectations of success, values of science and environmental attitudes. The survey items asked high school students to rank their agreement regarding 13 statements about their science motivation and environmental attitudes. The instrument utilized the 5-point Likert scale, which is a commonly used format in educational research. Students were asked to indicate their answers using the following rating scale: Strongly Agree = 5, Agree = 4, Neither Agree Nor Disagree = 3, Disagree = 2, and Strongly Disagree = 1. Students had access to the survey through the provided unique web link that each school received.

Qualtrics survey software was used for generating the survey and collecting responses. All students' responses were anonymous; no student name, school name or IP address was collected. Please see Appendix A (Study 1A) for survey's structure and items.

Participants and sample data. The survey was administered to 3,454 students from New Jersey, Pennsylvania and Connecticut. There were a total of 13 participating urban, suburban and private high schools. However, only 3,259 responses were considered complete and were utilized for further analysis. The data was collected in two phases. 1,911 responses from 11 schools were collected during phase one, which spanned May 2015 to June 2015. Phase two lasted from September 2015 to January 2016. In this phase 1,543 students from 3 schools took the survey.

The phase one data containing 1,764 valid responses was used for conducting exploratory factor analysis (EFA), whereas 1,495 valid responses from the phase two data set were utilized for performing confirmatory factor analysis (CFA). Since many validations of the survey instrument performed during the Rasch analysis are similar to those done by CFA, it made sense to use the same phase two data set with 1,495 student responses for the Rasch analysis.

Even though it is possible to perform the Rasch analysis on the incomplete data by using estimations of the structural parameters of the Rasch model (Verhelst & Glas, 1993), a listwise deletion (the complete removal of the respondent's answers to all items even if just one answer was missing) was implemented. This decision was based on two factors. First, the number of incomplete responses was small (56), meaning that removing them was not going to have a big impact on the analysis results. Second, while conducting the factor analysis (EFA and CFA), listwise deletion was used to remove incomplete answers. Hence, it was decided to use the same approach to missing data and remove incomplete answers by applying listwise deletion. After removing incomplete data, the 1,439 responses that were left were implemented for the Rasch analysis. This data set was identical to the one used during CFA analysis.

The Rasch analysis. Prior to conducting this analysis, the survey's data was analyzed by conducting both EFA and CFA. During the EFA, three different factor structures were considered: the two-factor, the three-factor and the four-factor models. These models were further analyzed using their fit to data, their alignment with the research constructs, and their factor loadings. As a result of these analyses, the three-factor model was chosen as the most appropriate factor representation of the SIEVEA's data. Then, the three-factor model was tested for reliability and validity using partial-confirmatory factor analysis (PCFA) and CFA. The three-factor model performed very well on all reliability and validity tests, thus confirming the correctness of the model selection made during the EFA. Consequently, the Rasch analysis of this study was conducted using three constructs/latent variables mapped to the three-factor model's factors.

The three-factor model contained the following factors: Factor #1 – "Science Identities and Motivation," Factor #2 – "Environmental Attitudes," Factor #3 – "Science Values." The model's factor structure with items comprising each factor is shown in Figure 2.15. Item #9's ("I can use technology for learning science content") data was removed from the analysis, since Study 1A showed that it was a problematic item and should be deleted from the survey.

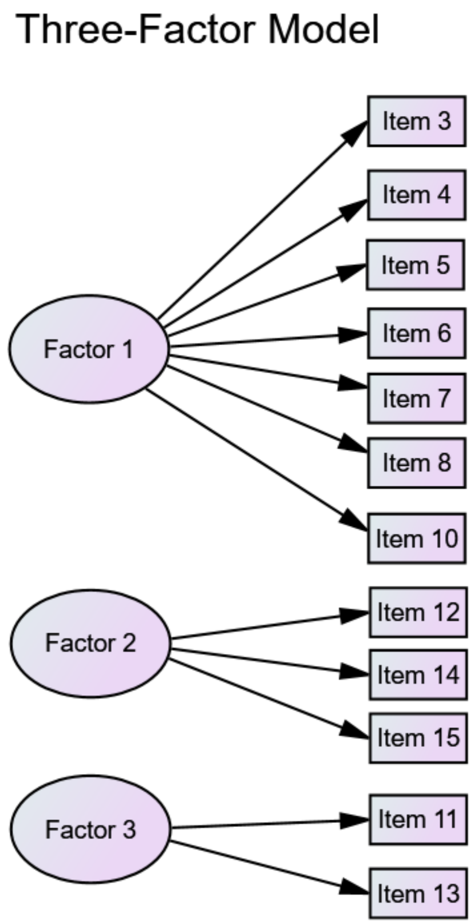


Figure 2.15. The Visual Representation of the Three-Factor Model

The three-factor model's structure was confirmed and validated during CFA, which helped test the a priori model, identify the "fit statistics," explain the covariation and the relationship among the observable and latent variables (Stevens, 1996). The CFA showed that the instrument's items could be grouped into three pools, with each one representing a single construct. This therefore makes it possible to carry out the Rasch analysis for each construct (Weller et al., 2013).

This study closely followed Sjaastad's (2014) guideline in order to do the Rasch analysis on the SIEVEA's data. Since the SIEVEA used the Likert scales and had more than two possible answer choices, the study utilized the polytomous Rasch model, not the dichotomous one. The polytomous model can be applied in two variations: the rating scale model and the partial credit model (Andrich, 1978). For the current study, the partial credit model was used, because it is more generic than the rating scale model and because it has the default calculation mode of most polytomous analysis tools.

There are multiple software packages that can be used to conduct the Rasch measurement analysis. These Rasch analysis software programs include FACETS (Linacre, 2014), RUMM 2030 (Andrich, Lyne, Sheridan, & Luo, 2011), WinSteps (Linacre, 2012), ConQuest2 (Wu, Adams, Wilson, & Haldane, 2007), and ConstructMap (Kennedy, Wilson, Draney, Tutunciyan, & Vorp, 2011). For assessing the SIEVEA instrument's quality, this study used the ConstructMap program, available free of charge on http://bearcenter.berkeley.edu/software/constructmap.

Separate Rasch analyses were performed on each subscale (construct) to ensure the unidimensionality assumption of the model is satisfied. Factor #3 ("Science Values") has only two items in the three-factor model. Unfortunately, no productive Rasch measurement calculations can be performed for a construct with one or two items (Gothwal et al., 2009). Because of this constraint, the Rasch analyses were performed for first two factors only: "Science Identities and Motivation" and "Environmental Attitudes."

During the Rasch analysis, the following calculations and analyses were performed: infit and outfit MSNQs were calculated for all items and participants; both items and participants' responses were analyzed for their fit; outlier participants were identified; item difficulties and their Thurstonian thresholds were calculated; item steps were mapped on a logit rating scale and Wright maps were constructed and analyzed to assess test targeting and discover person separation issues; ICC curves were plotted and scrutinized for any redundancies in response categories. Lastly, participants' raw ordinal scores were converted to interval scale and saved in preparation for conducting parametric statistical tests like t-test and ANOVA.

Since the constructs were discovered and confirmed during prior factor analyses and the Rasch analysis was performed on each construct separately, testing for the unidimensionality of data was not necessary. Additionally, measurement invariance testing during CFA established that the instrument was not biased for or against any participants based on gender. Because of this, performing DIF testing for gender during the Rasch analysis was redundant and, therefore, was omitted.

Results. ***Construct 1 (science identities and motivation).*** The calculated Cronbach's Alpha for this construct was 0.88. The polytomous Rasch measurement calculations indicated that items' infit mean-square fit statistics ranged from .69 to 1.14 (see Table 2.20). All these values fell well within the acceptable range of .5 – 1.5 (Linacre, 2012) indicating a good fit. Moreover, all values except one (item #6) were in more stringent range of .7 – 1.3 proposed by Bond and Fox (2001). The infit MNSQ of item #6 was .69; just a tad shy of the lower bound of .7 of Bond and Fox's suggested range. All this indicates a good, productive item fit meaning all items work well together in capturing their corresponding construct. Additionally, since all infit MNSQs fell within the narrow range of acceptable values, this showed one more time that the items satisfied the unidimensionality requirement (Eckes, 2005).

Table 2.18

Item Estimates (Construct C1) - Infit Mean Squares

Item #	MNSQ	0.75	0.83	0.92	1.00	1.08	1.17	1.25	1.33
Item 3	1.07					*			

Item 4	1.11	*
Item 5	.98	*
Item 6	.69	*
Item 7	1.03	*
Item 8	1.14	*
Item 10	1.00	*

Furthermore, there were only 61 participants (4% of the entire data set) with their normalized outfit outside of the range $-2 < t < 2$. This indicates that these participants' responses did not fit well within the Rasch model (meaning, their actual responses were significantly different from the responses predicted by the Rasch model). In order to assess the impact of these outliers on the model, these participants' responses were removed from the data set, following the guidance of Fortus and Vedder-Weiss (2014), and another run of the polytomous Rasch analysis was conducted. The results of this run showed no significant change in the model. There was a little change in all items' mean-square fit statistics. Some mean-square fit statistics did not change, whereas others changed by less than .04. The Cronbach's Alpha also changed very little: it increased from .88 to .89. Additionally, there was hardly any change in the Wright map and Thurstonian thresholds. Since removing these outliers did not have a significant impact on the model, it was decided to keep them in the data set. Indeed, there was no reason to believe that these responses were somewhat less reliable than the remaining responses.

Next, the items' difficulties and Thurstonian thresholds were calculated (see Table 2.21). Since these items are not used to measure an aptitude but rather a latent variable called "Science Identities and Motivation," their difficulties indicate how strongly the item expresses the latent variable in comparison to other items.

Table 2.19
Item Statistics (Construct C1) - Item Difficulty and Thurstonian Thresholds

		Thurstonian Thresholds			
Item #	Difficulty	1 → 2	2 → 3	3 → 4	4 → 5
3	-.67	-2.79	-1.64	-.24	1.98
4	-.88	-3.38	-1.98	-.59	2.45
5	-.55	-2.80	-1.55	-.13	2.30
6	.35	-2.11	-.41	1.09	2.84
7	-.20	-2.92	-1.58	.41	3.31
8	.34	-2.32	-.66	.82	3.51
10	1.60	-.91	.88	2.48	3.95

According to data in Table 2.21, item #10 ("My friends and family recognize me as a scientist") indicates the most extreme statement regarding to "Science Identities and Motivation" construct. On the other hand, item #4 ("I am confident I can master the skills taught in my science class") is the least extreme statement.

In Figures 2.16 and 2.17 items and their steps were placed on the C1 (Construct 1: Science Identities and Motivation) continuum using the logit scale. While locations of item difficulties (Figure 2.16) depict how items are ordered from the least difficult to the most difficult, the item step thresholds show how the instrument creates a "ruler" for measuring students' responses (see Figure 2.17).

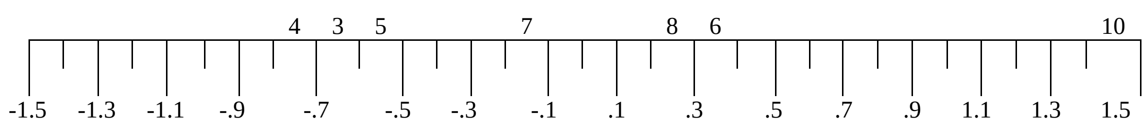


Figure 2.16. Locations of Item Difficulties on the C1 Continuum

On the item step thresholds "ruler" (Figure 2.17) each item has 4 steps ("markers") indicating transitioning from one response category to the next one. For example, item 3 has the following steps: 3.1, 3.2, 3.3, and 3.4. Here item.step 3.1 has the logit value of about -2.8 and indicates the transitioning from category 0 ("Strongly Disagree") to category 1 ("Disagree"). Likewise, item.step 3.2 has the logit value of about -1.6 and indicates the transitioning from "Disagree" to "Neither Agree Nor

Disagree”, item.step 3.3 is located at about -.2 logit value and corresponds to the switch from “Neither Agree Nor Disagree” to “Agree”, and item.step 3.4 is located at about 2 logit and specifies the shift from “Agree” to “Strongly Agree.”

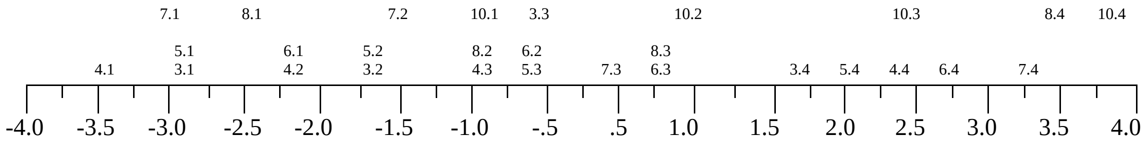


Figure 2.17. Locations of Item Step Thresholds on the C1 Continuum

The Wright map is another useful and convenient way for depicting item steps (their Thurstonian thresholds) on the logit scale so their relative difficulties can be visually inspected and compared. In addition, the Wright map shows how student responses were distributed against item steps (see Figure 2.18).

The Wright map was used to assess how well the instrument targeted the survey participants. Additionally, the Wright map was analyzed to find out how well the survey separated persons in regard to this construct and discover any floor or ceiling problems.

```
============================================================================
   raw      Students              Thurstonian Thresholds
----------------------------------------------------------------------------
    |28                XXXXX|
    |                       |
4   |27                  XXX|Item 10.4
    |                       |
    |26              XXXXXXX|Item 8.4
    |                       |Item 7.4
3   |25             XXXXXXXX|Item 6.4
    |24            XXXXXXXXX|
    |23            XXXXXXXXX|Item 4.4 Item 10.3
    |22           XXXXXXXXXX|Item 5.4
2   |                       |Item 3.4
    |21       XXXXXXXXXXXXXX|
    |20     XXXXXXXXXXXXXXXX|
    |19   XXXXXXXXXXXXXXXXXX|Item 6.3
1   |18 XXXXXXXXXXXXXXXXXXXX|Item 8.3 Item 10.2
    |17 XXXXXXXXXXXXXXXXXXXX|
    |16  XXXXXXXXXXXXXXXXXXX|Item 7.3
0   |15    XXXXXXXXXXXXXXXXX|
    |14          XXXXXXXXXXX|Item 3.3 Item 5.3
    |13         XXXXXXXXXXXX|Item 6.2
    |12          XXXXXXXXXXX|Item 4.3 Item 8.2
-1  |11           XXXXXXXXXX|Item 10.1
    |         XXXXXXXXXXXXXX|
    |8                 XXXXX|
    |7                    XX|Item 3.2 Item 5.2 Item 7.2
-2  |6                    XX|Item 4.2
    |5                    XX|Item 6.1
    |4                    XX|Item 8.1
    |3                    XX|Item 3.1 Item 5.1
-3  |2                     X|Item 7.1
    |                       |
    |1                     X|Item 4.1
    |                       |
-4  |                       |
    |                       |
    |0                    XX|
============================================================================
          Each X represents 6 students, each row is 0.255 logits
```

Figure 2.18. Wright Map for Thurstonian Thresholds (Construct 1)

As can be seen from the Wright Map (see Figure 2.18), there were some ceiling and floor problems with the instrument. 25 students scored higher (raw score of 28, logit score of about 4.5) than the highest item.step threshold (logit 3.95; item 10, step 4). Therefore, the instrument could not differentiate between these 25 high scorers. Likewise, there are 7 students with minimum raw score of 0 (logit score of about -4.4). But this floor issue is not as significant as the ceiling issue since it affects only a small number of respondents.

Also, both the “Locations of Item Step Thresholds on the C1 Continuum” figure and the Wright diagram (see Figures 2.17 and 2.18) show some item redundancies. Items 3 (“Learning science in school will help me to succeed later in life”) and 5 (“I consider science topics very interesting and engaging”) have very similar step thresholds. This can also be detected by inspecting their thresholds values in Table 2.21. Indeed, the absolute values of differences of these two items’ step thresholds’ logit vales are rather small: step 1 – .1 logit; step 2 – .09 logit; step 3 – .11 logit; step 4 – .32 logit. This shows that one of these items can be removed from the instrument without decreasing its ability to separate respondents.

In addition, the Wright diagram showed that the instrument had some skewness towards respondents with stronger ability. Indeed, more respondents were located at the top part of the diagram (above logit 0) than below.

Item Characteristics Curves (ICC) for construct 1 (see Figure 2.19) were used to evaluate the appropriateness of item response categories and discover any redundancies. For each item the ICC contains 5 curves: one curve per response category. The points of curve intersections indicate the transition between corresponding categories. Therefore, for an item with appropriate response categories these points must follow the same order as the categories: their x-axis values should increase as curves progress from 0 to 4 going from left to right. By visually inspecting ICCs for all 7 items it became clear that all response categories are appropriate. Additionally, the curves showed that there were no redundancies. Indeed, any redundancy would make points of intersections to flip (reverse their order). As can be seen from the graphs in Figure 2.19, no flipping of order happened for these items.

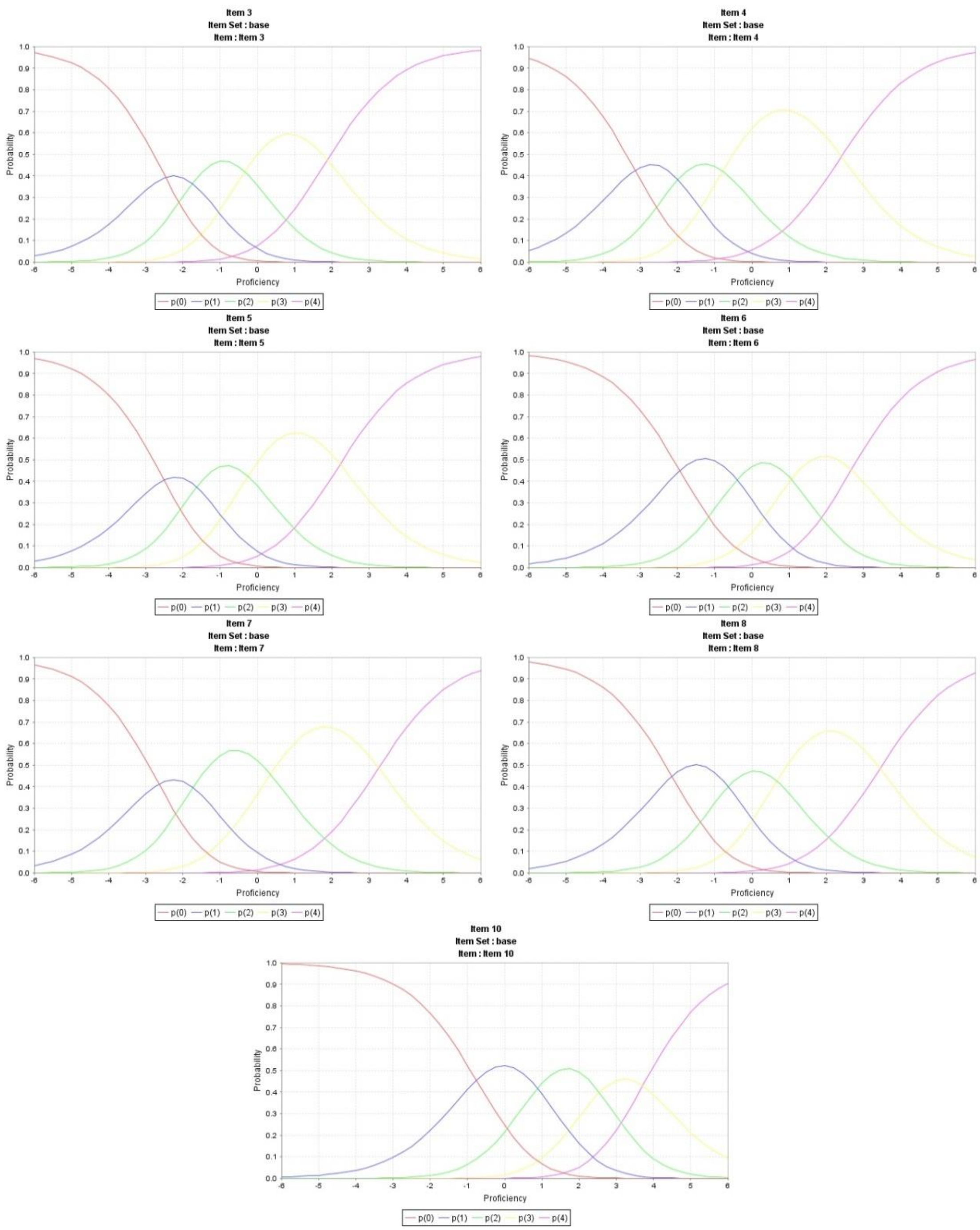


Figure 2.19. Item Characteristic Curves (Construct 1)

Construct 2 (environmental attitudes). The calculated Cronbach's Alpha for this construct was 0.74. Infit mean-square statistics for all 3 items were calculated as part of the polytomous Rasch measurement calculations. Their values ranged from .91 to 1.15 (see Table 2.22) indicating a very good fit based on 7 - 1.3 range proposed by Bond and Fox (2001). Therefore, it can be concluded that all three items representing "Environmental Attitudes" fit well and do good job in capturing the construct. Moreover, these infit MNSQ values provided another evidence of the unidimensionality of data (Eckes, 2005).

Table 2.20
Item Estimates (Construct C2) - Infit Mean Squares

Item #	MNSQ	0.75	0.83	0.92	1.00	1.08	1.17	1.25	1.33
Item 12	1.04					*			
Item 14	1.15						*		
Item 15	.91			*					

Out of 1,439 participants 160 (about 11% of the entire data set) had normalized outfit outside of the range -2 < t < 2 meaning that these participants' responses were outliers. Again, these outliers were removed from the data set and the polytomous Rasch analysis was run on the resulting data set so the outliers' impact can be measured. The results showed some differences: items 12 and 14 MNSQ fit statistics decreased by .05, whereas item 3 MNSQ increased by .05, the Cronbach's Alpha increased from .74 to .81. There also was a slight change in the Wright map and Thurstonian thresholds: item.step 12.4 and 15.4 switched their order on the logit scale. However, there was no reason to believe that these outlier responses were not reliable and, since these changes did not significantly alter the overall model, it was decided to keep them in the data set.

The item difficulties and items' Thurstonian thresholds for "Environmental Attitude" are listed in Table 2.23. Item 15 ("I am interested in reading website, articles or watching TV programs, documentary movies about the environmental issues") indicated the strongest environmental attitude, whereas item 14 ("It is important for all people to be engaged in vital environmental issues") was the weakest attitude. Item 12 ("I would like to become more active on important environmental issues") had .01 logit as difficulty sitting difficulty-wise exactly in the middle of items 14 and 15.

Table 2.21
Item Statistics (Construct C2) - Item Difficulty and Thurstonian Thresholds

		Thurstonian Thresholds			
Item #	Difficulty	1 → 2	2 → 3	3 → 4	4 → 5
12	.01	-2.70	-1.31	.75	3.28
14	-.39	-2.66	-1.70	.10	2.68
15	.39	-2.50	-.44	1.04	3.44

Item difficulties measured in logit units were visualized on the C2 (Construct 2: Environmental Attitude) continuum (see Figure 2.20). Likewise, Figures 2.21 showed how item steps work together as a single measurement tool for measuring this construct. This "ruler" presented a good, almost evenly distributed item step thresholds covering the entire -4 to 4 logit range. This indicated that the instrument should be doing a good job in test targeting.

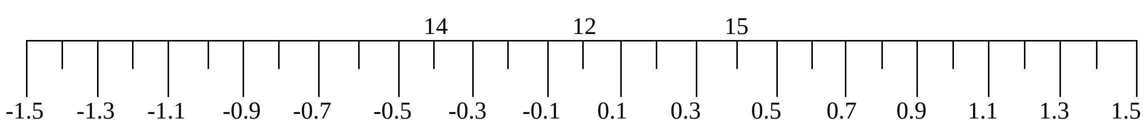


Figure 2.20. Locations of Items' Difficulties on the C2 Continuum

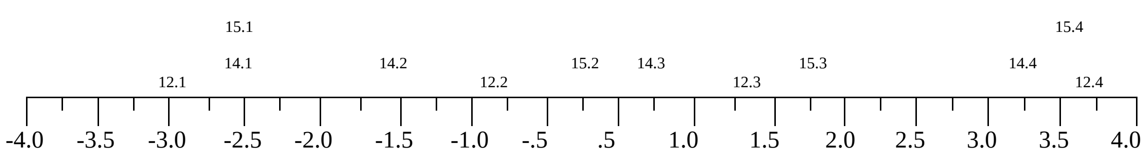


Figure 2.21. Locations of Item Steps' Thresholds on the C2 Continuum

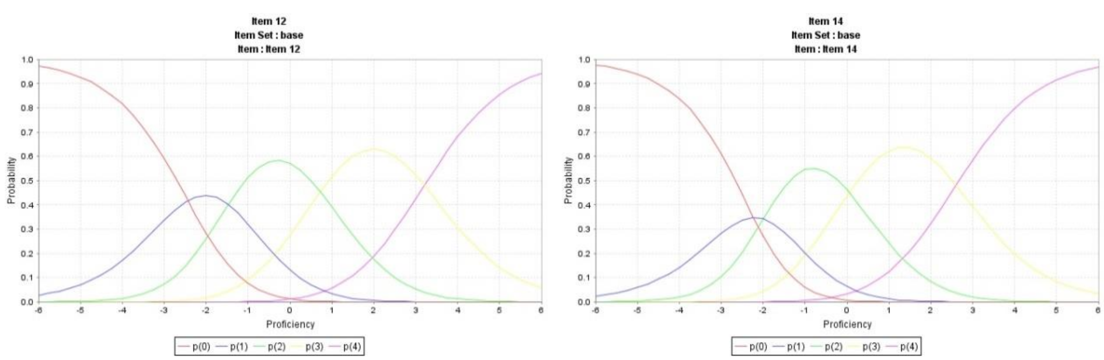
The Wright map (see Figure 2.22) confirmed that the instrument properly targets the pool of respondents. There is no ceiling issue and no significant floor issue. This means that the instruments covered almost all persons based on their environmental attitude. Also, there was no skewing in data distribution: the responses were almost symmetrically distributed with respect to 0 logit score.

```
============================================================================
   raw      Students                Thurstonian Thresholds
----------------------------------------------------------------------------
4   |                       |
    |                       |
    |12                 XXXX|Item 15.4
    |                       |Item 12.4
3   |                       |
    |11                XXXXX|Item 14.4
    |                       |
    |10             XXXXXXXX|
2   |                       |
    |9      XXXXXXXXXXXXXXXX|
    |                       |
    |8    XXXXXXXXXXXXXXXXXX|Item 15.3
1   |                       |
    |   --------------------|Item 12.3
    |7  XXXXXXXXXXXXXXXXXXXX|
0   |                       |Item 14.3
    |6     XXXXXXXXXXXXXXXXX|
    |                       |Item 15.2
    |5          XXXXXXXXXXXX|
-1  |4               XXXXXXX|
    |                       |
    |3                  XXXX|Item 12.2
    |                       |Item 14.2
-2  |2                    XX|
    |                       |
    |1                     X|Item 15.1
    |                       |Item 12.1 Item 14.1
-3  |                       |
    |0                    XX|
    |                       |
    |                       |
-4  |                       |
============================================================================
          Each X represents 13 students, each row is 0.255 logits
```

Figure 2.22. Wright Map for Thurstonian Thresholds (Construct 2)

The Wright map showed one area of improvement: the instrument did not have enough markers on the "ruler" within the middle range of logit values. Indeed, the area between logit values -1 and 2.25 has only 4 markers (12.3, 14.3, 15.2, and 15.3), whereas lots of student scores are located within this range. In order to improve the instrument in this regard, it will make sense to add more environmental attitude measuring items to the questionnaire.

Visual examination of ICCs for all 3 items showed no issues with the appropriateness of item response categories (see Figure 2.23). All transitions points from one category to the next one followed the expected order on the x-axis. Moreover, there were no redundant response categories (reverse ordering of points of intersections). However, it should be noted that for item 14 transition points 1 -> 2 and 2 -> 3 between categories were close on the x-axis. This means that there were no significant proficiency differences between respondents who answered "Disagree" on item 14 and respondents who answered "Neither Agree Nor Disagree" on the same item. This could be just another manifestation of the mid-category issue (DeMars & Erwin, 2004; Kulas et al., 2008).



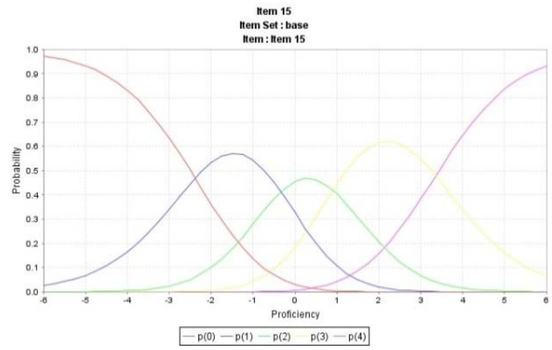


Figure 2.23. Item Characteristic Curves (Construct 2)

Discussion. The SIEVEA instrument was already validated during the CFA analysis using various statistical tests and model fit indexes. However, it was important to do the Rasch analysis of the SIEVEA data since the Rasch analysis provided many additional methods, which were not available in the CFA analysis, for assessing the instrument's reliability and validity. As expected, the Rasch analysis strengthened the case of the SIEVEA instrument's reliability and validity by making it possible to examine the instrument from a different angle. The Rasch analysis allowed looking at item and person fits in order to assess how well the individual items and person responses fit within the item pool and the response pool respectively. It also uncovered any issues with item and person misfits, including outlier responses.

Overall, all items comprising the "Science Identities and Motivation" and "Environmental Attitudes" constructs of the instrument had a good fit. Their mean squared statistics were well within the acceptable range of values indicating that the items worked well as a group and demonstrating once more the unidimensional nature of the data for each construct. The data contained some outlier responses: about 4% responses for the "Science Identity and Motivation" construct and 11% responses for the

“Environmental Attitudes” construct. However, removing these outliers from the data did not have a significant impact on the results of the Rasch analysis.

Additionally, the Rasch analysis helped in evaluating the test targeting. In other words, it assisted in answering the following questions: 1) How well did this particular test/instrument work for this particular pool of respondents? 2) Was the instrument able to separate different respondents based on their ability level? 3) Were there any ceiling or floor issues, namely, was the instrument able to properly assess respondents with very high or very low ability levels?

It turned out that SIEVEA did an acceptable job in targeting the respondent pool. The item steps covered the entire range of logit values, providing appropriate measuring marks for evaluating the respondents’ abilities. However, some areas of improvement were noticed. The instrument can be enhanced by adding more items for the “Environmental Attitudes” construct, since the existing item steps did not generate enough marks to provide sufficient granularity in scoring the respondents with abilities within the middle region of the logit scale. The "Science Value" did not have enough items in order to conduct productive Rasch analysis. Therefore, on top of adding more items to the instrument for the “Environmental Attitudes” construct, it is necessary to expand the instrument by adding more items for measuring the "Science Value" construct.

Next, the Rasch analysis involved looking at all items and response categories to assess whether there were any redundancies in items and categories. All response categories were adequate and without any redundancies. However, although the instrument did a good job in measuring the "Science Identity and Motivation" construct, it can be improved by removing some redundant items. Namely, items 3 and 5 had very close step thresholds, indicating a possibility of dropping one of these items without decreasing the instrument's usefulness and reliability.

Also, there was a ceiling issue with measuring of the “Science Identities and Motivation” construct, although it was not significant. It may be worth adding one more item expressing a more extreme statement on this construct than the current items in order to improve the instrument's ability to measure the respondents with very high ability (25 respondents). Only 7 respondents had lower ability levels than the threshold value of the instrument's lowest category (4.1 – item 4, category 1). Therefore, this floor issue can be ignored.

Conclusion. In conclusion, the Rasch analysis allowed for looking at the SIEVEA instrument from a different perspective. It performed in-depth analyses on SIEVEA data and once again confirmed the instrument's reliability and validity for measuring the “Science Identities and Motivation” and “Environmental Attitudes” constructs.

The Rasch analysis also uncovered some issues with the instrument and pointed to areas of potential enhancements, like removing or adding items in order to increase the instruments' test targeting. Furthermore, the Rasch measurement calculations assigned a new, logit-based score to each participant for each construct in lieu of their raw score. Converting raw, ordinal scores to logit, interval scale scores was an important preliminary step for preparing the survey's data to be used in various parametric statistical tests like t-test and ANOVA.

References

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, *43*(4), 561-573.

Andrich, D. (1988). *Rasch Models for Measurement*: Thousand Oaks, CA: Sage.

Andrich, D., Lyne, A., Sheridan, B., & Luo, G. (2011). *RUMM 2030.* Perth: RUMM Laboratory. Also available at http://www.rummlab.com.au/

Brentani, E., & Golia, S. (2007). Unidimensionality in the Rasch model: How to detect and interpret. *Statistica*, *67*(3), 253-261.

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). Mahwah, NJ: Erlbaum.

Boone, W. J., & Scantlebury, K. (2006). The role of Rasch analysis when conducting science education research utilizing multiple-choice tests. *Science Education*, *90*(2), 253-269.

Cappelleri, J. C., Lundy, J. J., & Hays, R. D. (2014). Overview of classical test theory and item response theory for the quantitative assessment of items in developing patient-reported outcomes measures. *Clinical Therapeutics, 36*(5), 648-662.

DeMars, C. E., & Erwin, T. D. (2004). Scoring neutral or unsure on an identity development instrument for higher education. *Research in Higher Education*, *45*(1), 83-95.

Ding, L. (2012, February). Applying Rasch theory to evaluate the construct validity of brief electricity and magnetism assessment. In *2011 Physics Education Research Conference* (Vol. 1413, No. 1, pp. 175-178). AIP Publishing.

Eckes, T. (2005). Examining rater effects in TestDaF writing and speaking performance assessments: A many-facet Rasch analysis. *Language Assessment Quarterly: An International Journal*, *2*(3), 197-221.

Gothwal, V. K., Wright, T. A., Lamoureux, E. L., & Pesudovs, K. (2009). Rasch analysis of the quality of life and vision function questionnaire. *Optometry & Vision Science*, *86*(7), 836-844.

Hambleton, R., Swaminathan, H., Rogers, H. (1991). *Fundamentals of Item Response Theory*. Newbury Park, California: Sage Publications, Inc.

Hibbard, J. H., Mahoney, E. R., Stockard, J., & Tusler, M. (2005). Development and testing of a short form of the patient activation measure. *Health Services Research*, *40*(6p1), 1918-1930.

Kennedy, C., Wilson, M. R., Draney, K., Tutunciyan, S., & Vorp, R. (2011). *ConstructMap Downloads.* Retrieved from http://bearcenter.berkeley.edu/ConstructMap/download.php.

Kulas, J. T., Stachowski, A. A., & Haynes, B. A. (2008). Middle response functioning in Likert-responses to personality items. *Journal of Business and Psychology*, *22*(3), 251-259.

Lee, J. (2004). Evaluating the effectiveness of instructional resource allocation and use: IRT and HLM analysis of NAEP teacher survey and student assessment data. *Studies in Educational Evaluation*, *30*(2), 175-199.

Linacre, J. M. (2012). *Winsteps ministep. Rasch-model computer program.* Retrieved from http://www.winsteps.com/winman/.

Linacre, J. M. (2014). *Facets Rasch measurement computer program* (Version 3.71.4)[Computer software]. Chicago, IL: Winsteps.com.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. New York, NY: Routledge.

Mallinson, T. (2007). Why measurement matters for measuring patient vision outcomes. *Optometry & Vision Science*, *84*(8), 675-682.

Mallinson, T., Stelmack, J., & Velozo, C. (2004). A comparison of the separation ratio and coefficient α in the creation of minimum item sets. *Medical Care*, *42*(1), 1-17.

McCrum-Gardner, E. (2008). Which is the correct statistical test to use? *British Journal of Oral and Maxillofacial Surgery*, *46*(1), 38-41.

Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*, *104*(1), 1-15.

Rasch, G. (1960). *Probabilistic models for some intelligence and achievement tests.* Copenhagen: Danish Institute for Educational Research.

Rasch, G. (1980). *Probabilistic Models for Intelligence and Attainment Tests.* Chicago, IL: University of Chicago Press.

Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. *Psychological Bulletin*, *114*(3), 552.

Schulz, W., & Fraillon, J. (2011). The analysis of measurement equivalence in international studies using the Rasch model. *Educational Research and Evaluation*, *17*(6), 447-464.

Sjaastad, J. (2014). Enhancing measurement in science education research through Rasch analysis: Rationale and properties. *Nordic Studies in Science Education*, *10*(2), 212-230.

Stenner, A. J., Burdick, H., Sanford, E. E., & Burdick, D. S. (2005). How accurate are Lexile text measures? *Journal of Applied Measurement*, *7*(3), 307-322.

Stevens, J. (1996). *Applied multivariate statistics for the social sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Tennant, A., & Conaghan, P. G. (2007). The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis Care & Research*, 57(8), 1358-1362.

Verhelst, N. D., & Glas, C. A. (1993). A dynamic generalization of the Rasch model. *Psychometrika*, *58*(3), 395-415.

Wang, W. C., Yao, G., Tsai, Y. J., Wang, J. D., & Hsieh, C. L. (2006). Validating, improving reliability, and estimating correlation of the four subscales in the WHOQOL-BREF using multidimensional Rasch analysis. *Quality of Life Research*, *15*(4), 607-620.

Weller, J. A., Dieckmann, N. F., Tusler, M., Mertz, C. K., Burns, W. J., & Peters, E. (2013). Development and testing of an abbreviated numeracy scale: A Rasch analysis approach. *Journal of Behavioral Decision Making*, *26*(2), 198-212.

Wilson, M. (2005). *Constructing measures: An item response modeling approach.* Mahwah, NJ: Lawrence Erlbaum Associates.

Wright, B. D. (1992). Raw scores are not linear measures: Rasch vs. Classical Test Theory CTT Comparison. *Rasch Measurement Transactions*, *6*(1), 208. Retrieved from http://www.rasch.org/rmt/rmt61n.htm

Wright, B. D. (1997). A history of social science measurement. *Educational Measurement: Issues and Practice*, *16*(4), 33-45.

Wright, B. D., Linacre, J. M., Gustafson, J. E., & Martin-Lof, P. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, *8*(3), 370.

Wright, B. D., & Masters, G. N. (1982). *Rating Scale Analysis. Rasch Measurement.* Chicago, IL: MESA Press.

Wright, B. D., & Stone, M. H. (1979). *Best Test Design. Rasch Measurement*: Chicago, IL: Measurement, Evaluation, Statistics, and Assessment Press.

Wu, M. L., Adams, R. J., Wilson, M. R., & Haldane, S. (2007). *ACER ConQuest Version 2.0*. Manual Retrieved from http://www.rasch.org/rmt/rmt133o.htm.

Zain, A. N. M., Samsudin, M. A., & Jusoh, A. (2010). Using the Rasch Model to Measure Students' Attitudes toward Science in" Low Performing" Secondary Schools in Malaysia. *International Education Studies*, *3*(2), 56-63.

Chapter 3. Study 2: Investigating High School Students' Science Identities and Motivation in Science, Values of Science and Environmental Attitudes

Introduction. The purpose of this study was to measure and analyze three constructs that influence students' science learning: students' science identities and motivation in science, students' attitudes toward the environment, and students' values of science. In order to research these constructs, quantitative research can be conducted using data collected in the form of motivation and attitude surveys (Shields & Rangarajan, 2013). For this study, the data collected via the SIEVEA survey instrument was used.

It was anticipated that the study would uncover statistically significant correlations between the three research constructs for the whole data set (Aschbacher, Li & Roth, 2010). Additionally, this study explored interests of high school students in different science subjects. Subsequently, the study looked into the distributions of students' science preferences and examined whether there were any correlations between students' science subject preferences and their science identities and motivation in science, values of science and attitudes toward the environment. It was predicted that students' science identities and motivation in science, values of science and attitudes toward the environment would vary depending on students' science subject preferences.

Studies conducted about two decades ago claimed that boys considered math and science topics useless, whereas girls raised concerns about "not being enough smart" and not doing well in math and science (AAUW, 1992). Likewise, Stark and Gray's (1999) study, conducted about two decades ago, utilized a national survey to measure students' preferences of certain science topics and found similar results regarding gender preferences in science learning. It turned out that boys and girls had different science subject preferences which changed as they moved from elementary to secondary school. Therefore, one of the goals of this study was to understand whether students' science subject preferences varied by gender. Consequently, the distributions of students' science preferences by gender were analyzed.

Previous studies established the existence of a gender effect on students' science attitudes, namely that boys had more positive attitudes toward science than girls (Weinburgh, 1995). In addition to attitudes towards science, gender was shown to influence science achievement, with females scoring lower than males (Schibeci & Riley, 1986). Additionally, several past studies regarding students' competence beliefs and values supported the idea that gender differences existed in various domains. For example, the Jacobs, Lanza, Osgood, Eccles, & Wigfield (2002) longitudinal studies established that gender differences were domain specific, rather than global, but that they did not systematically increase with age. Also, there were gender-specific differences in students' attitudes in science especially in the secondary schools (Stark & Gray, 1999). Therefore, it was important to examine recent data regarding high school students' science views and attitudes, and to determine whether the gender gap discovered by prior research still exists currently as it did two decades ago. Also, the current study explored whether any existing gender differences were similar to or different from past differences. It was hypothesized that male and female students would report science identities, motivational constructs and environmental attitudes differently.

Another area of the research involved the differences between urban and suburban students in regard to three research constructs and students' science preferences. Hence, this study examined the distributions of students' science preferences by participating school types (urban versus suburban). Lastly, past research highlighted that urban minority boys academically performed poorer than their female peers; in other words, gender differences existed in urban schools (Graham, 2000; Weaver-Hightower, 2003). Hence, this study looked into the differences between girls and boys across different school settings (urban vs. suburban) in order to uncover whether gender differences in high school students' science identities and motivation in science, values of science and environmental attitudes vary depending on the school type.

Here is the complete list of this study's research questions:

1. How do students' science subject preferences influence their science identities and motivation in science?
2. How do students' science subject preferences influence their values of science?
3. How do students' science subject preferences influence their attitudes toward the environment?
4. Are there any correlations between students' science identities and motivation in science, values of science and attitudes toward the environment?
5. Are there any gender differences in high school students' science identities and motivation in science, values of science and attitudes toward the environment?
6. Do urban and suburban high school students view their science identities and motivation in science, values of science and attitudes toward the environment differently?

7. Do gender differences in high school students' science identities and motivation in science, values of science and environmental attitudes vary depending on school type (urban vs. suburban)?
8. Do gender differences in high school students' science identities and motivation in science, values of science and environmental attitudes vary depending on students' science subject preferences (physical science vs. life science)?
9. Do students' science subject preferences vary by gender?
10. Do students' science subject preferences vary by school type: urban versus suburban?

Method. ***Participants.*** A total of 13 districts agreed to participate in this study. Data was collected in three different states: New Jersey, Pennsylvania and Connecticut. The sample included student participants from seven suburban schools, five urban schools and one private school. In total, 3,454 high school students took the survey. However, only 3,099 responses were complete and used for data analysis. In order to encourage students and districts to participate in this study, no ethnic, racial or other demographic information was collected, aside from gender. Since the survey was online, study participants had a choice of taking it either at home or in school. The survey participation was voluntary and anonymous; no IP addresses were collected. The detailed information about the study participants is shown in Table 3.1.

Table 3.1
Study Participants: By School, School Type and Gender

		Participants			
School	School Type	Total	Female	Male	Unknown Gender
S1	Suburban	166	90	76	0
S2	Urban	231	115	116	0
S3	Urban	425	218	205	2

S4	Urban	64	27	36	1
S5	Suburban	27	18	9	0
S6	Suburban	143	74	69	0
S7	Suburban	55	32	22	1
S8	Private	41	0	41	0
S9	Urban	105	63	42	0
S10	Suburban	165	97	67	1
S11	Suburban	46	30	16	0
S12	Suburban	1,364	750	611	3
S13	Urban	267	154	113	0

Note. Data shows participants with complete answers.

Data sources. The survey instrument SIEVEA was administered by using the Qualtrics website (rutgers.qualtrics.com). There were a total of 15 questions in the survey. The first survey item allowed the researchers to collect student gender data. The second question provided information with respect to participants' interest toward science subjects. Lastly, the remaining thirteen questions covered four survey constructs. These 13 questions were listed in no particular order. The survey utilized a 5-point Likert scale to capture student answers to those 13 questions. The answer choices were: strongly agree, agree, neither agree nor disagree, disagree and strongly disagree.

Since survey participants were high school students, the simplicity of the design was a priority. Among other reasons, the Likert-type scale was chosen so the survey format resembled a multiple-choice format test familiar to students. In addition to predefined, multiple choice questions, the second item provided a message box, in which students could type their favorite science subject if it was not present in the provided list. All questions used simple words and straightforward sentence structures. Although richer data could have been collected, the survey was restricted to a small number of questions, because of concern that students might become tired and fail to complete the survey or provide accurate answers.

Exploratory and confirmatory factor analyses were conducted prior to this study, in order to discover and confirm the survey's factor structure and validate the survey. Likewise, Cronbach's reliability analyses were conducted, which established the survey's reliability. Additionally, two constructs, science identities and motivation and environmental attitudes, were further validated using the Rasch analysis.

Results. The results of data analyses for all research questions follow.

Research Question 1: How do students' science subject preferences influence their science identities and motivation in science?

A one-way analysis of variance indicated statistically significant differences between science identity and motivation scores of students with different subject preferences. Table 3.2 shows the results of this test: F = 80.308 and p < .000.

Table 3.2
A One-Way ANOVA Summary
The Effects of Student Favorite Science Subject on their Science Identities and Motivation in Science

Source	*df*	*SS*	*MS*	*F*	*p*
Between-group	7	918.693	131.242	80.308	.000
Within-group	3,090	5,049.760	1.634		
Total	3,097	5,968.453			

Table 3.3 shows descriptive statistics (count, mean, standard deviation, and 95% confidence intervals) by favorite science subject. According to this data, the students who chose biology, physics and chemistry as their favorite science subjects held stronger science identities (M > 1) than the students who preferred Earth science, environmental science, forensics or other science subjects. The students with preference in physics had the strongest science identity and motivation (M = 1.251), whereas the students who favored Earth science had the lowest score (M = .440). The post hoc comparisons, using

the Tukey's procedure, confirmed these observations by producing statistically significant (p < .000) mean differences between these groups.

Table 3.3
Descriptive Statistics of Students' Science Identities and Motivation in Science By Favorite Science Subject
Means, Confidence Intervals and Standard Deviations

Favorite Subject	*N*	*M*	*SD*	*95% Confidence Interval*	
				Lower Bound	*Upper Bound*
Biology	817	1.091	1.283	1.003	1.179
Chemistry	810	1.037	1.298	.948	1.127
Other	265	.816	1.383	.649	.984
Earth Science	163	.440	1.103	.269	.610
Environmental Science	135	.516	1.293	.296	.736
Forensics	297	.573	1.292	.426	.721
Physics	258	1.251	1.301	1.092	1.411
None	353	-.616	1.177	-.740	-.493

Research Question 2: How do students' science subject preferences influence their values of science?

In order to answer this research question, a one-way analysis of variance test was conducted on students' values of science scores. The scores used for this test were the factor scores generated during factor analysis, not Rasch scores. The test produced statistically significant results (F = 4.633, p < .000) as shown in Table 3.4.

Table 3.4
A One-Way ANOVA Summary
The Effects of Student Favorite Science Subject on their Values of Science

Source	*df*	*SS*	*MS*	*F*	*p*
Between-group	7	21.075	3.011	4.633	.000
Within-group	3,090	2,008.169	.650		
Total	3,097	2,029.243			

Note. This test used students' factor scores.

According to descriptive statistics (see Table 3.5), the students who chose chemistry (M = .062) and physics (M = .051) as their favorite subjects valued science the most.

Table 3.5
Descriptive Statistics of Students' Values of Science
By Favorite Science Subject
Means, Confidence Intervals and Standard Deviations

				95% Confidence Interval	
Favorite Subject	*N*	*M*	*SD*	*Lower Bound*	*Upper Bound*
Biology	817	.017	.812	-.039	.073
Chemistry	810	.062	.794	.007	.116
Other	265	-.096	.834	-.197	.005
Earth Science	163	.011	.791	-.112	.133
Environmental Science	135	.024	.824	-.116	.164
Forensics	297	.046	.770	-.042	.133
Physics	258	.051	.828	-.051	.152
None	353	-.199	.812	-.284	-.114

Note. Descriptive statistics used students' factor scores.

Research Question 3: How do students' science subject preferences influence their attitudes toward environment?

Table 3.6 shows the results of a one-way ANOVA test conducted to answer this research question. As indicated by this test, students' attitudes toward the environment varied due to their subject preferences with statistical significance: F = 29.155 and p < .000.

Table 3.6
A One-Way ANOVA Summary
The Effects of Student Favorite Science Subject on their Attitudes toward Environment

Source	*df*	*SS*	*MS*	*F*	*p*
Between-group	7	310.484	44.355	29.155	.000
Within-group	3,090	4,700.912	1.521		
Total	3,097	5,011.395			

Table 3.7 lists various descriptive statistics of students' attitudes toward the environment by their favorite science subject. According to this data, the students who chose environmental science, Earth science and biology as their favorite subjects held stronger attitudes toward the environment than the students with different science subject preferences. The students with a preference for environmental science held the strongest environmental attitudes (M = 1.410). The Tukey's post hoc comparisons confirmed these findings by producing statistically significant ($p < .000$) mean differences between these groups.

Table 3.7
Descriptive Statistics of Students' Attitudes toward Environment
By Favorite Science Subject
Means, Confidence Intervals and Standard Deviations

				95% Confidence Interval	
Favorite Subject	*N*	*M*	*SD*	*Lower Bound*	*Upper Bound*
Biology	817	.914	1.219	.830	.997
Chemistry	810	.641	1.203	.558	.724
Other	265	.766	1.250	.615	.917
Earth Science	163	.905	1.211	.718	1.092
Environmental Science	135	1.410	1.288	1.191	1.629
Forensics	297	.658	1.229	.517	.798
Physics	258	.769	1.257	.615	.923
None	353	-.039	1.296	-.174	.097

Research Question 4: Are there any correlations between students' science identities and motivation in science, values of science and attitudes toward the environment?

In order to answer this research question, Pearson correlation coefficients between all the constructs were calculated. Table 3.8 explains the details of Pearson's correlation among three constructs.

Table 3.8
Intercorrelations for Three Constructs

	1	2	3
1. Science Identities and Motivation	–		
2. Environmental Attitudes	.096**	–	
3. Science Values	.104**	.102**	–

**. Correlation is significant at the .01 level (2-tailed).

Even though the correlation coefficients were not high (all three coefficients were around .1), they all were significant ($p < .01$), indicating statistically significant relationships between all three constructs.

Research Question 5: Are there any gender differences in high school students' science identities and motivation in science, values of science and attitudes toward the environment?

In order to answer this research question, a one-way analysis of variance test was conducted. The test indicated statistically significant differences by gender in students' science identities and motivation ($F = 29.610$, $p < .000$) and values of science ($F = 12.240$, $p < .000$). On the other hand, the differences in students' environmental attitudes were not statistically significant ($p = .457$). The results of this test are shown in Table 3.9.

Table 3.9
A One-Way ANOVA Summary
Gender Differences - All Three Constructs

Construct	*Source*	*df*	*SS*	*MS*	*F*	*p*
C1	Between-group	1	56.559	56.559	29.610	.000
	Within-group	3,089	5,900.427	1.910		
	Total	3,090	5,956.986			
C2	Between-group	1	.897	.897	.555	.457
	Within-group	3,089	4,997.282	1.618		
	Total	3,090	4,998.179			
C3	Between-group	1	7.982	7.982	12.240	.000
	Within-group	3,089	2,014.482	.652		

	Total	3,090	2,022.465

Note. C1 = Science Identities and Motivation, C2 = Environmental Attitudes, C3 = Science Values. C1 and C2 scores are Rasch scores. C3 scores are factor scores.

Constructs scores' comparison by gender showed that males ($M = .909$) had stronger science identities and higher expectations of success in science than their female peers ($M = .637$). However, females ($M = .047$) assign higher value to science than males do ($M = -.055$). See Table 3.10 for details.

Table 3.10
Descriptive Statistics of Three Constructs by Gender
Means, Confidence Intervals and Standard Deviations

					95% Confidence Interval	
Construct		*N*	*M*	*SD*	*Lower Bound*	*Upper Bound*
C1	Female	1,668	.637	1.372	.572	.703
	Male	1,423	.909	1.393	.836	.981
C2	Female	1,668	.721	1.277	.660	.782
	Male	1,423	.687	1.267	.621	.753
C3	Female	1,668	.047	.809	.008	.086
	Male	1,423	-.055	.806	-.097	-.013

Note. C1 = Science Identities and Motivation, C2 = Environmental Attitudes, C3 = Science Values. C1 and C2 scores are Rasch scores. C3 scores are factor scores.

Research Question 6: Do urban and suburban high school students view their science identities and motivation in science, values of science and attitudes toward environment differently?

A one-way analysis of variance test (see Table 3.11) showed no statistically significant differences between urban and suburban students regarding their science identities and motivation ($p = .461$) and values of science ($p = .222$). However, the test indicated a statistically significant difference between urban and suburban students' attitudes towards environment ($F = 16.254$, $p < .000$).

Table 3.11

A One-Way ANOVA Summary
School Type Differences - All Three Constructs

Construct	*Source*	*df*	*SS*	*MS*	*F*	*p*
C1	Between-group	1	1.053	1.053	.544	.461
	Within-group	3,056	5,915.269	1.936		
	Total	3,057	5,916.322			
C2	Between-group	1	26.202	26.202	16.254	.000
	Within-group	3,056	4,926.443	1.612		
	Total	3,057	4,952.645			
C3	Between-group	1	.978	.978	1.493	.222
	Within-group	3,056	2,001.316	.655		
	Total	3,057	2,002.294			

Note. C1 = Science Identities and Motivation, C2 = Environmental Attitudes, C3 = Science Values. C1 and C2 scores are Rasch scores. C3 scores are factor scores.

Urban students' environmental attitude scores had a mean value of .830 which was higher than the mean score (M = .637) of suburban students (see Table 3.12).

Table 3.12
Descriptive Statistics of Three Constructs by School Type
Means, Confidence Intervals and Standard Deviations

					95% Confidence Interval	
Construct		*N*	*M*	*SD*	*Lower Bound*	*Upper Bound*
C1	Urban	1,092	.734	1.278	.658	.810
	Suburban	1,966	.772	1.450	.708	.837
C2	Urban	1,092	.830	1.278	.755	.906
	Suburban	1,966	.637	1.265	.581	.693
C3	Urban	1,092	.024	.799	-.024	.071
	Suburban	1,966	-.014	.815	-.050	.022

Note. C1 = Science Identities and Motivation, C2 = Environmental Attitudes, C3 = Science Values. C1 and C2 scores are Rasch scores. C3 scores are factor scores.

Research Question 7: Do gender differences in high school students' science identities and motivation in science, values of science and environmental attitudes vary depending on the school type (urban vs. suburban)?

According to the research question 5, gender differences were statistically significant for construct 1 (science identities and motivation) and construct 3 (value of science). Likewise, the research question 6 showed that school type differences were statistically significant for the construct 3 (environmental attitudes). Therefore, in order to answer this research question, two-way analysis of variance tests were conducted for all three constructs using gender and school type as independent variables. Table 3.13 contains the results of this test for construct 1.

Table 3.13
A Two-Way ANOVA Summary
Gender Differences by School Type
Construct 1: Science Identities and Motivation in Science

Source	*df*	*SS*	*MS*	*F*	*p*
Between-group	1	52.842	52.842	27.520	.000
Interaction (Gender*School Type)	2	2.020	1.010	.526	.591
Within-group	3,046	5,848.668	1.920		
Total	3,050	7,656.383			

Since the p-value of the interaction component (Gender*School Type) was not statistically significant (p = .591), it followed that gender differences in students' science identities and motivation do not vary by school type. Table 3.14 contains descriptive statistics of construct 1 by gender and school type.

Table 3.14
Descriptive Statistics by Gender and School Type
Construct 1: Science Identities and Motivation in Science

Gender	*School Type*	*N*	*M*	*SD*
Female	Urban	577	.590	1.290
	Suburban	1,091	.663	1.414
	Total	1,668	.637	1.372
Male	Urban	512	.896	1.248
	Suburban	870	.907	1.484
	Total	1,382	.903	1.401
Total	Urban	1,089	.734	1.279

Suburban	1,961	.771	1.450
Total	3,050	.758	1.392

Mean scores of both females and males in suburban schools were slightly higher than those in urban schools: females (M = .663 suburban; M = .590 urban), males (M = .907 suburban; M = .896 urban). As Figure 3.1 shows, there is some interaction between gender and school type: two lines corresponding to each school type are not parallel, because the difference between suburban and urban female students' mean scores is larger than the difference between suburban and urban male students' mean scores. However, this difference is not statistically significant.

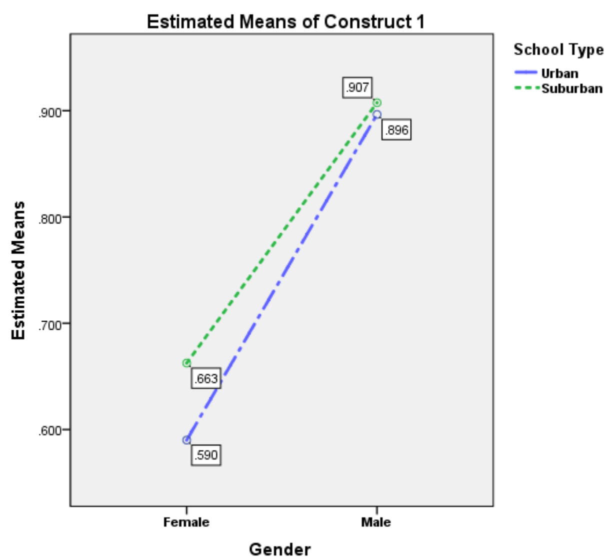


Figure 3.1. Means of Construct 1 by Gender and School Type

Likewise, the two-way analysis of variance test for construct 3 showed no statistically significant interaction between gender and school type for construct 3 (see Table 3.15). The p-value of the interaction component (Gender*School Type) was .207.

Table 3.15
A Two-Way ANOVA Summary
Gender Differences by School Type
Construct 3: Science Values

Source	*df*	*SS*	*MS*	*F*	*p*
Between-group	1	9.146	9.146	14.034	.000
Interaction (Gender*School Type)	2	2.052	1.026	1.575	.207
Within-group	3,046	1,984.980	.652		
Total	3,050	1,995.285			

As data in Table 3.16 shows, female and male students in urban schools had higher mean scores that those in suburban schools. The mean scores of females were .095 in urban and .022 in suburban schools (the scores are factor scores calculated during factor analysis). Likewise, the mean scores of males were -.052 in urban and -.061 in suburban schools.

Table 3.16
Descriptive Statistics by Gender and School Type
Construct 3: Science Values

Gender	*School Type*	*N*	*M*	*SD*
Female	Urban	577	.095	.808
	Suburban	1,091	.022	.808
	Total	1,668	.047	.809
Male	Urban	512	-.052	.779
	Suburban	870	-.061	.822
	Total	1,382	-.057	.806
Total	Urban	1,089	.026	.797
	Suburban	1,961	-.015	.815
	Total	3,050	.000	.809

Note. C3 = Science Values. C3 scores are factor scores.

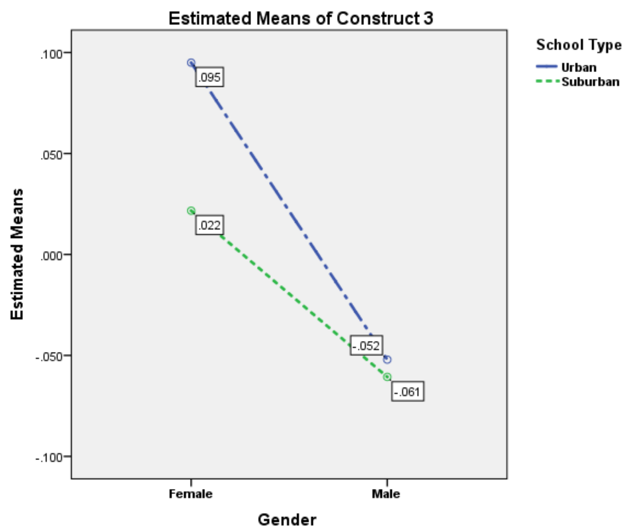


Figure 3.2. Means of Construct 3 by Gender and School Type

Figure 3.2 shows that there was some non-trivial interaction between gender and school type: the two lines corresponding to each school type are not parallel. However, the interaction was not statistically significant.

Lastly, the test for the environmental attitude construct by school type and gender also showed no significant interaction with p = .288 (see Table 3.17).

Table 3.17
A Two-Way ANOVA Summary
School Type Differences by Gender
Construct 2: Environmental Attitudes

Source	*df*	*SS*	*MS*	*F*	*p*
Between-group	1	28.610	28.610	17.755	.000
Interaction (School Type *Gender)	2	4.014	2.007	1.245	.288
Within-group	3,046	4,908.281	1.611		
Total	3,050	4,939.394			

Research Question 8: Do gender differences in high school students' science identities and motivation in science, values of science and environmental attitudes vary depending on students' science subject preference (physical science vs. life science)?

In order to investigate this research question, students' favorite science subjects were grouped into two categories: physical science and life science. Physics, chemistry and Earth science fell under the physical science category, whereas biology and environmental science were combined into the life science category. Students with science preferences that could not be characterized as physical or life sciences (e.g., forensics) were excluded from consideration. This resulted in a smaller data set of 2,177 responses.

Analysis of the research question 5 showed that gender differences were statistically significant only for the construct 1 (science identities and motivation) and the construct 3 (value of science). Because of this, this research question was considered only for these two constructs. A two-way ANOVA test was conducted to examine the effect of gender and science subject preference (physical science vs. life science) on students' science identities and motivation in science and values of science.

The test results for construct 1 showed a statistically significant interaction ($F(2, 2173) = 5.576$, $p = .004$) between the effects of gender and science subject preference on students' science identities and motivation in science (see Table 3.18). This indicates a significant interaction between gender and science subject preference regarding students' science identities and motivation in science.

Table 3.18
A Two-Way ANOVA Summary
Gender Differences by Science Type
Construct 1: Science Identities and Motivation in Science

Source	*df*	*SS*	*MS*	*F*	*p*
Corrected Model	3	45.995	15.332	9.217	.000
Intercept	1	2,035.053	2,035.053	1,223.439	.000
Gender	1	21.595	21.595	12.983	.000
Interaction(Gender*Science Type)	2	18.551	9.275	5.576	.004

Error	2,173	3,614.542	1.663
Total	2,177	5,867.477	
Corrected Total	2,176	3,660.537	

Table 3.19 shows the descriptive statistics of construct 1 by gender and science type. Overall, males had a higher mean score (M = 1.125) on construct 1 than females (M = .900), which was the entirely expected result, because the research question 5 produced similar results. Further examination of the data demonstrated that the difference between males and females with preference in physical sciences was much more pronounced than the difference between males and females with preference in life sciences. Indeed, among students with physical science preferences, males had a mean score of 1.175, whereas females had a mean score of .786, with a difference of .389. On the other hand, the mean scores of males and females with life science preferences were much closer: males had a mean score of 1.026 and females had a mean score of 1.004. This resulted in a rather small difference: .022.

Table 3.19
Descriptive Statistics by Gender and Science Type
Construct 1: Science Identities and Motivation in Science

Gender	*Science Type*	*N*	*M*	*SD*
Female	Life Science	596	1.004	1.267
	Physical Science	545	.786	1.269
	Total	1,141	.900	1.272
Male	Life Science	350	1.026	1.356
	Physical Science	686	1.175	1.291
	Total	1,036	1.125	1.314
Total	Life Science	946	1.012	1.300
	Physical Science	1,231	1.003	1.295
	Total	2,177	1.007	1.297

These results show that males' scores were higher than females' scores due primarily to students with preferences for physical sciences. Figure 3.3 depicts the interaction between students' gender and their preferred science type.

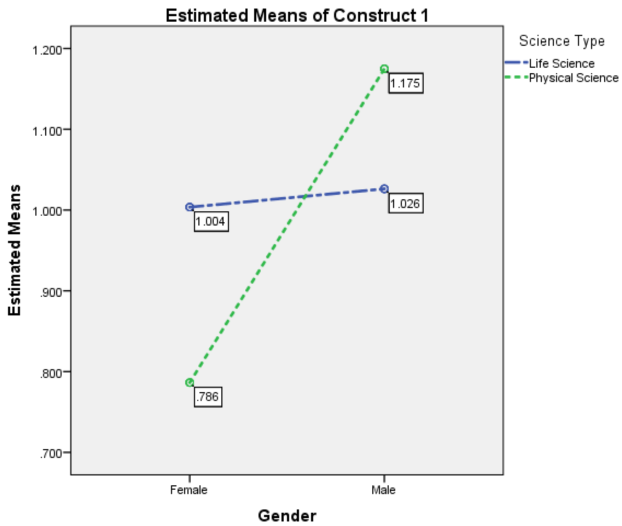


Figure 3.3. Means of Construct 1 by Gender and Science Type

As for construct 3, the two-way analysis of variance test indicated no statistically significant interaction between students' gender and their preferred science type (see Table 3.20). The p-value of the interaction component (Gender*Science Type) was .303.

Table 3.20
A Two-Way ANOVA Summary
Gender Differences by Science Type
Construct 3: Science Values

Source	*df*	*SS*	*MS*	*F*	*p*
Corrected Model	3	5.88	1.960	3.028	.028
Intercept	1	1.924	1.924	2.973	.085
Gender	1	5.308	5.308	8.200	.004
Interaction(Gender*Science Type)	2	1.548	.774	1.196	.303
Error	2,173	1,406.466	.647		
Total	2,177	1,415.556			
Corrected Total	2,176	1,412.346			

Research Question 9: Do students' science subject preferences vary by gender?

In order to answer this research question, a Chi-square test was conducted. The test results indicated statistically significant differences in female and male science subject preferences. The value of Pearson Chi-Square was 125.815 (df = 7), which was significant with $p < .000$. Table 3.21 shows the breakdown of students' science subject preferences by gender. In addition to counts in each category (for example, 518 females prefer biology), the table also displays percentages of each category within gender, favorite science subject and total.

Table 3.21
Students' Science Subject Preferences by Gender

		Favorite Science Subject								
Gender		Biology	Chemistry	Other	Earth Science	Environmental Science	Forensics	Physics	None	Total
Female	Count	518	395	130	73	78	192	77	205	1,668
	% within Gender	31.1%	23.7%	7.8%	4.4%	4.7%	11.5%	4.6%	12.3%	100%
	% within Favorite Science Subject	63.9%	48.8%	49.4%	44.8%	57.8%	64.6%	29.8%	58.1%	54%
	% of Total	16.8%	12.8%	4.2%	2.4%	2.5%	6.2%	2.5%	6.6%	54%
Male	Count	293	415	133	90	57	105	181	148	1,422
	% within Gender	20.6%	29.2%	9.4%	6.3%	4.0%	7.4%	12.7%	10.4%	100%
	% within Favorite Science Subject	36.1%	51.2%	50.6%	55.2%	42.2%	35.4%	70.2%	41.9%	46%
	% of Total	9.5%	13.4%	4.3%	2.9%	1.8%	3.4%	5.9%	4.8%	46%
Total	Count	811	810	263	163	135	297	258	353	3,090
	% within Gender	26.2%	26.2%	8.5%	5.3%	4.4%	9.6%	8.3%	11.4%	100%
	% within Favorite Science Subject	100%	100%	100%	100%	100%	100%	100%	100%	100%

According to data, biology is the most preferred subject for females followed by chemistry and forensics. On the other hand, males like chemistry the most, whereas biology is their second choice and physics is third. Earth science and environmental science are not popular choices for both females and males.

Research Question 10: Do students' science subject preferences vary by school type: urban versus suburban?

Chi-square test indicated that there were statistically significant differences between urban and suburban students' science subject preferences. The Pearson Chi-Square was 26.675 (df = 7) with p < .000. Students' science subject preferences by school type are summarized in Table 3.22.

Table 3.22

Students' Science Subject Preferences by School Type

		Favorite Science Subject								
School Type		Biology	Chemistry	Other	Earth Science	Environmental Science	Forensics	Physics	None	Total
Urban	Count	261	252	113	59	50	125	97	135	1,092
	% within School Type	23.9%	23.1%	10.3%	5.4%	4.6%	11.4%	8.9%	12.4%	100%
	% within Favorite Science Subject	32.2%	31.4%	43.0%	36.6%	37.0%	43.0%	39.8%	38.6%	35.7%
	% of Total	8.5%	8.2%	3.7%	1.9%	1.6%	4.1%	3.2%	4.4%	35.7%
Suburban	Count	549	551	150	102	85	166	147	215	1,965
	% within School Type	27.9%	28.0%	7.6%	5.2%	4.3%	8.4%	7.5%	10.9%	100%
	% within Favorite Science Subject	67.8%	68.6%	57.0%	63.4%	63.0%	57.0%	60.2%	61.4%	64.3%
	% of Total	18.0%	18.0%	4.9%	3.3%	2.8%	5.4%	4.8%	7.0%	64.3%
Total	Count	810	803	263	161	135	291	244	350	3,057
	% within School Type	26.5%	26.3%	8.6%	5.3%	4.4%	9.5%	8.0%	11.4%	100%
	% within Favorite Science Subject	100%	100%	100%	100%	100%	100%	100%	100%	100%

Even though the differences by school type were statistically significant, they were less pronounced than differences by gender. For example, biology and chemistry were the topmost popular science subjects in both urban and suburban schools, with approximately equal popularity (23.9% biology and 23.1% chemistry in urban schools; 27.9% biology and 28.0% chemistry in suburban schools). Their popularity in suburban schools was about 4% higher than their popularity in urban schools. Conversely, forensics was more popular with urban students (11.4%) than with suburban ones (8.4%). Preferences in Earth science and environmental science were only slightly dissimilar, with less than a .5% difference.

Discussion. The data collected using the SIEVEA instrument was rich and very useful. It allowed for conducting multiple statistical tests in order to answer the research questions of this study.

According to the results, students' science subject preferences significantly influence their science identities and motivation in science, values of science and environmental attitudes. Moreover, students with preferences in biology, chemistry and physics have strong science identities and motivation in science, whereas students who prefer chemistry and physics value science the most. Additionally, students whose favorite subject is environmental science have the strongest environmental attitudes. This outcome was intuitively predictable and appeared very reasonable. Students, who like challenging science subjects like physics or chemistry, should be both interested in science and highly motivated to learn science. Similarly, strong environmental attitudes should go along with a keen interest in learning about environment. Still, it is significant that these intuitive knowledge was confirmed by statistical analysis of a large, heterogeneous response data.

The prior study, which conducted an exploratory factor analysis on the SIEVEA survey instrument (Study 1A), indicated correlations between the motivational and environmental attitude items of the SIEVEA despite evidence of discriminant validity, meaning the correlation coefficients between environmental attitude items and motivational items were significantly less than those between environmental attitude items. As this study clearly demonstrated, all three research constructs were correlated. It should be noted that the correlations between students' science identities and motivation in science, values of science and attitudes toward the environment were not high; all calculated Pearson correlation coefficients were around .1. However, they all were statistically significant, indicating non-accidental interrelations among constructs.

Next, this study uncovered several important gender-related differences. First, it turned out that males' and females' science identities and motivation in science varied significantly. Specifically, males had stronger science identities than females. Also, males were more motivated in science learning than females were. These gender differences in science identities and motivation in science were primarily among students with preferences for physical sciences like physics, chemistry and Earth science. Indeed, there was a statistically significant interaction between gender and science subject preference regarding students' science identities and motivation in science.

Weinburgh (1995) observed similar differences by gender in students' attitudes toward science. Specifically, that study established that males had better attitudes toward science than females. Likewise, Schibeci & Riley (1986) examined how students' background and perceptions influenced their science attitude and achievement. Their study used five independent variables: sex, race, home environment, amount of homework, and parents' education. The dependent variables were student perception of science instruction, student attitudes, and student achievement. The study determined that sex, race, and the home environment had considerable influence on student achievement in science. Additionally, it found that students' attitude influenced their achievement.

Interestingly, females ascribed higher value to science than males, whereas there was no difference between males' and females' environmental attitudes. Wigfield and Eccles (1992) observed that when individuals found their tasks difficult to accomplish, they were more likely to value the task. Since this study showed that females scored higher on science value than males, it will be interesting to explore how females and males rate the difficulty of science related tasks.

Schibeci (1984) uncovered that science subject preferences varied by gender. According to his study's results, physics and chemistry were boys' favorite subjects, whereas girls had a clear preference for biology. This study reached similar conclusions. It confirmed that females and males had statistically significant differences in their science subject preferences. Additionally, it elaborated that females liked biology the most, followed by chemistry and forensics and males liked chemistry the most, followed by biology as their second choice and physics as third.

Investigation of differences between urban and suburban students revealed a rather interesting outcome: there were no statistically significant variances in urban and suburban students' science identities and motivation and in how they value science. However, urban students had better environmental attitudes than suburban students. Urban students' positive environmental attitudes were not unexpected, since the Stern, Powell & Ardoin's (2011) study also showed that urban students had positive attitudes regarding environmental responsibility. These results were orthogonal to those produced by analysis of gender differences. Due to this peculiarity, combined effects of the gender and school type (urban vs suburban) were not different from the individual effects of each variable. Indeed, there were no interaction effects between gender and school type with respect to these three research constructs.

Lastly, there were statistically significant differences between urban and suburban students' science subject preferences. Even though both urban and suburban students had a strong preference for biology and chemistry, forensics was much more popular with urban students than with suburban students.

Conclusion. This study demonstrated the effectiveness of the SIEVEA instrument. The data collected by this survey was used to investigate students' science identities, motivation in science, values of science and environmental attitudes. It allowed for conducting a rich, quantitative research using various statistical tests and for exploring how students' gender, school type, and favorite science subject affect the research constructs. The effects of students' attributes on the constructs were analyzed both independently and collectively in order to discover significant interactions.

Since sustainability is now an integral part of Next Generation Science Standards (NGSS, 2014), the scholars felt the need to question how effectively it is being taught in schools (Feinstein & Kirchgasler, 2015). They expressed doubts as to whether the political aspect of sustainability should and could be taught in environmental science classes. As such, they recommended collaboration between the science and social studies teachers in teaching students about sustainability challenges. The SIEVEA instrument allowed for the examining of high school students' views regarding the environment and environmental issues in general.

The study's results indicated that students' science subject preferences influence their science identities, environmental attitudes and how they value science. Also, there were statistically significant differences between males and females with respect to science identities and motivation in science and science value, and between urban and suburban students regarding their environmental attitudes. Additionally, it turned out that students' science subject preferences significantly vary by gender and by school type (urban versus suburban). These results can be used by both educational researchers and practitioners for developing instructional and teaching strategies, which can facilitate development of stronger science identities, increase students' motivation in science learning and improve their environmental attitudes.

Since the SIEVEA survey was designed for high school students, it can be used to conduct longitudinal studies in the future. For example, the researchers can use the SIEVEA instrument with the same subjects once a year for 4 consecutive years. By collecting data from the same students in the 9th to 12th grades, it will be possible to explore how their science identities, motivation in science and environmental attitudes develop and transform while they go through high school. It will be interesting to find out if any noticeable and significant changes are taking place during the students' high school years. Another recommendation for future research is to conduct the SIEVEA survey twice a year: once at the beginning and once at the end of the year (Jovanovic & King, 1998). This will allow for determination of how students' science identities, motivation and environmental attitudes evolve as they progress through an academic year.

References

American Association of University Women Educational Foundation. (1992). *How schools shortchange girls: Executive summary.* Washington, DC: Author.

Aschbacher, P. R., Li, E., & Roth, E. J. (2010). Is science me? High school students' identities, participation and aspirations in science, engineering, and medicine. *Journal of Research in Science Teaching, 47*(5), 564-582.

Carlone, H. B., & Johnson, A. (2007). Understanding the science experiences of successful women of color: Science identity as an analytic lens. *Journal of Research in Science Teaching, 44*(8), 1187-1218.

De Groot, J. I., & Steg, L. (2010). Relationships between value orientations, self-determined motivational types and pro-environmental behavioural intentions. *Journal of Environmental Psychology, 30*(4), 368-378.

Eccles, J. S., Wigfield, A., & Schiefele, U. (1998). Motivation to succeed. In W. Damon (Series Ed.) & N. Eisenberg (Vol. Ed.), *Handbook of Child Psychology: Vol. 3. Social, emotional and personality development* (5th ed., pp. 1017–1094). New York, NY: Wiley.

Feinstein, N. W., & Kirchgasler, K. L. (2015). Sustainability in science education? How the Next Generation Science Standards approach sustainability, and why it matters. *Science Education, 99*(1), 121-144.

Graham, S. (2000). *Academic achievement and ethnic minority students: Constraints and opportunities.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Jacobs, J. E., Lanza, S., Osgood, D. W., Eccles, J. S., & Wigfield, A. (2002). Changes in children’s self-competence and values: Gender and domain differences across grades one through twelve. *Child Development, 73*(2), 509-527.

Jovanovic, J., & King, S. S. (1998). Boys and Girls in the Performance-Based Science classroom: Who’s doing the performing? *American Educational Research Journal, 35*(3), 477-496.

Schibeci, R. A. (1984). Attitudes to science: An update. *Studies in Science Education, 11*, 26-59.

Schibeci, R. A., & Riley, J. P. (1986). Influence of students' background and perceptions on science attitudes and achievement. *Journal of Research in Science Teaching, 23*(3), 177-187.

Shields, P. M., & Rangarajan, N. (2013). *A playbook for research methods: Integrating conceptual frameworks and project management.* Stillwater, OK: New Forums Press.

Stark, R., & Gray, D. (1999). Gender preferences in learning science. *International Journal of Science Education*, *21*(6), 633-643.

Stern, J. M., Powell, B. R., & Ardoin, M. N. (2011). Evaluating a constructivist and culturally responsive approach to environmental education for diverse audiences. *The Journal of Environmental Education*, *42*(2), 109-122.

Weaver-Hightower, M. (2003). The “boy turn” in research on gender and education. Review of *Educational Research*, *73*(4), 471-498.

Weinburgh, M. (1995). Gender differences in student attitudes toward science: A meta-analysis of the literature from 1970 to 1991. *Journal of Research in Science Teaching*, *32*(4), 387-398.

Wigfield, A., & Eccles, J. S. (1992). The development of achievement task values: A theoretical analysis. *Developmental Review*, *12*(3), 265-310.

Chapter 4. Study 3: Do I see myself as a science person? A reflection on urban youth's science identities during the River City project

Introduction. Urban students are often portrayed as unengaged, underachieving students who are incapable of mastering various subjects, including science (Kane, 2012), and as "at risk" students who eventually fail (Bryan & Atwater, 2002). This stereotypical link between low socioeconomic status students and underachievement was brought up by various studies and statistics (Steele, 1997). Some researchers have put blame for academic failure on the students, their parents and their communities (Rubin, 2007). Others contend that, since low socioeconomic status students tend to encounter adults who have low academic achievement, hold either low paid or non-professional jobs (Oyserman, Bybee & Terry, 2006), they view themselves as one of them.

Student identities get shaped by what they think and how they feel about their "self" (James, 1890) and by the ways in which students think of themselves as learners (Kane, 2012). The construct of identity refers to "a person's understandings of self in relationship to the world, including people, places, events, material objects, and semiotic systems" (Kane, 2012, p. 458). Schools have a profound impact on shaping student identities, by creating either a notion of "smartness" or a lack of it (Hatt-Echevarria, 2005). This notion is especially applicable to the learning sciences, where students' science identity gets shaped based on how they view themselves and how others (friends, teachers, and parents) perceive them as science learners. The researcher of this study, who is also a high school science teacher, encountered several cases where students gave up learning science and claimed that they were not good at science and that they constantly failed science since their elementary school years.

Students' continuous success was linked to the development of their identities (Horn, 2008). According to Carlone and Johnson (2007), recognizing oneself and being recognized by others as a "science person" gives the person a strong science identity. For this reason, Andersen and Ward (2014) placed strong emphasis on the improvement of students' own perceptions regarding their science identities. Likewise, Kane (2012) found a critical connection between science content learning and the construction of science identities, especially in African American children. Similar studies done in the mathematics content area found connections between mathematical identity development and students' participation in math classes (Gresalfi, 2004; Sfard & Prusak, 2005). Studies done in the past with high school students revealed that students moderately applied their school science knowledge to socioscientific discussions (Ratcliffe, 1997); however, socioscientific contexts provide excellent opportunities for students, allowing them to provide justifications based on their real-life experiences (Osborne, Erduran, & Simon, 2004). More recent studies with high school students demonstrate similar patterns during socioscientific discussions: instead of providing scientific evidence, students provided more instances of economical, environmental and ethical support (Simon & Amos, 2011).

Past research showed that many minority students did not think that math and science were valuable for them, due to the lack of examples of successful science people among their friends, families or in their communities; indeed, many of these students considered science to be a subculture of White, male, Western culture (Hines, 2003). This may explain why only 5% of scientists holding doctorate degrees are African American or Hispanic (National Science Foundation, 2001). This type of a mindset and predisposition to failure inhibits student motivation and engagement in science classrooms, which inevitably leads to high failure (Anagnostopoulos, 2003) and low retention rates in urban schools (Swanson, 2004).

In this study, urban youth's science identities were examined as they progressed through various activities in an online collaborative learning environment called the River City. This research measured how urban students' science identities evolved throughout the study when students situated themselves and developed a different view about themselves in the communities of practice (Lave & Wenger, 1991). It was well documented by the prominent cognitive anthropologist Jean Lave and educational theorist Etienne Wenger, that community of practice cultivates motivation and develops identity (Lave & Wenger, 1991; Wenger, 1998). In the context of the River City project, this statement is translated into an assumption that students brought their pre-existing scientific knowledge, motivation in science and science identities to the project.

According to studies conducted in the past, students are capable of generating thoughtful questions in a science context; moreover, question asking leads to critical reasoning (Chin & Osborne, 2008). Thought-provoking and probing questions help in fostering higher-order thinking skills that are essential for students to succeed in the 21st century (Saavedra & Opfer, 2012; Schwartz & Fischer, 2006). Therefore, a small amount of time (2 days) of the River City project was devoted to students' question crafting.

Additionally, there is a critical connection between students' science identities and their participation in science classes and learning of science content (Kane, 2012). During the current study, students worked on real-world problems in an online collaborative learning environment, which simulated genuine professional activities.

Literature review. ***The River City.*** The River City project was initially designed by Harvard professors and is currently managed by ActiveWorlds Incorporated. The River City's (http://rivercity5.activeworlds. com) multi-user virtual environment (MUVE) is great for teaching students scientific inquiry and 21st century skills. It is based on the National Science Standards (National Research Council, 1996), aligns well with the Next Generation Science Standards (NRC, 2013) and features environmental and epidemiological topics. Likewise, almost all practices of science were integrated in the River City project, which allowed students to understand what science is exactly about (Duschl & Grandy, 2013; Osborne, 2014).

In this project, students acted as River City visitors. As River City visitors, students traveled back in time and brought their 21st century knowledge and technology to address 19th century problems. River City is a town plagued with health problems, and students worked together in groups to investigate the root causes of these problems. They collected data, formed hypotheses, developed controlled experiments for testing their hypotheses, and made recommendations based on their findings. Here is how the authors, who designed and created River City, describe the project. *"The River City unit is based on students collaboratively investigating a virtual "world" consisting of a city with a river running through it, different forms of terrain that influence water runoff, houses, industries, and institutions such as a hospital and university. The learners themselves populate the city, along with non-player characters (NPCs), digital objects that can include audio or video clips, and computer-based agents. River City contains over fifty digital objects from the Smithsonian's collection, plus 'data collection stations' that* *provide detailed information about water samples at various spots in the world."* (Dede, Ketelhut, & Ruess, 2003, p.3).

One of the advantages of the MUVE is the ability of multiple users, acting as computer-based avatars, to collaborate in a virtual environment at the same time (Dede, Ketelhut, & Ruess, 2003). Students navigate through this virtual world as avatars, virtual persons who can interact with other avatars, communicate and share ideas with town's residents. All communication among avatars and the computer-based agents takes place either through text chat or virtual gestures (Clarke & Dede, 2009). Although avatars and residents cannot utter any sounds, there are some audio clips embedded in the project, such as mosquitoes buzzing, people coughing, sounds of migrating birds, and audio-effects when using teleportation to move within the town or from one season to another. The River City project includes historical facts and pictures of historical characters and events taken from the Smithsonian Institute (Clarke & Dede, 2009).

Figure 4.1. The River City User Interface

The user interface of the River City is engaging and simple to use (see Figure 4.1). Previous findings showed that the software's interface, its user-friendliness, and its familiarity to students play a crucial role in ensuring an effective learning process (Curtis & Lawson, 2001). Although the River City has an attractive interface and is user-friendly, students' answers to the pre-test questionnaire indicated that students were not familiar with game-based learning environments, such as the River City. Therefore, it was not clear at the beginning how students would respond to learning science through a web-based imaginary world and avatars. One of the reasons that this study utilized the River City for data collection was the idea that students considered the River City project as a strategy video game, where they had to come up with the solution for deciphering the problems taking place in the River City.

Even though the River City project is appropriate for middle and high school students spanning grades five through twelve, most of the research in this project was conducted with middle school students in classroom-based settings. Moreover, most of the time, it was carried out with disengaged, non-motivated middle school students (Dede, Ketelhut, & Ruess, 2003). Furthermore, unlike previously conducted studies, the current study explored upper-grade students' science identities in online, rather than classroom, environments. As it was mentioned earlier, students' questioning skills were assessed and evaluated through the project's embedded resources. Students received a special training on questioning skills in Day 3 and Day 4. The intervention was based on Chin's (2006) questioning framework, including a) modeling higher-order thinking questioning asking, b) providing question stems, c) providing question taxonomy (Bloom's revised taxonomy), and d) asking students to practice questioning skills by posing and answering their questions on Google Docs. Data generated during students' online collaboration within the River City MUVE and the Google Docs postings was hand-coded and qualitatively analyzed. Both the River City and Google Docs allowed students to create and submit electronic messages, which should help in increasing peer engagement as hypothesized by Peters and Hewitt (2010).

Research established that when the schoolwork was aligned with students' personal interests, they became engaged in the learning process (Assor, Kaplan, & Roth, 2002). The world has faced grave issues, such as the Ebola virus outbreak and the Zika virus epidemic, in recent times. Students witnessed the seriousness of those emergent diseases and the global concern about them. Since the problems plaguing River City were similar to these real-world events, it was hypothesized that they would ignite interest among the students and intrigue them to unravel the root causes of these problems.

Critical thinking. As Daniel Pink has written in his book, “Problems need solutions. That's elementary. But sometimes smart solutions need a few more problems” (Pink, 2015, p. 153). Very often, daily problems are the “leftovers” of already solved problems. Furthermore, the “smart strategy” will be proposing solutions based on the previous problem, rather than by inventing alternative solutions for tackling each new problem (Needham & Begg, 1991).

Although various studies intertwined critical thinking, problem solving, higher-order thinking, reasoning in one continuum, Lewis & Smith (1993) differentiated them based on philosophical and psychological perspectives of their nature. According to them, the logical reasoning and perfections of thinking are in the realm of the philosophers, whereas the thinking process leading to meaningful construction is of the psychologists' interest. Furthermore, research showed that knowledge construction, along with peer interaction and active inquiry learning, enhances science education for K-12 students (Pearson, Moje, & Greenleaf, 2010). Facione (1990) described critical thinking as a process that involves analyses, inferences, evaluation, deductive and inductive reasoning. Much earlier, Watson & Glasser (1980) gave a similar definition of critical thinking, but they added interpretation and recognition of assumptions as important components of critical thinking.

Even though activities requiring lower-order thinking skills are more abundant in school curricula than the ones with higher-order thinking skills (Saavedra & Opfer, 2012), many schools have started to adapt and design a curriculum that emphasizes higher-order thinking skills. Indeed, it has been more than a decade since many schools started integrating good thinking in their curriculum, which led to learning environments that welcome and value good thinking skills and where students can practice those important skills (Venville, Adey, Larkin, Robertson, & Fulham, 2003). Moreover, students need to do the thinking themselves in order to improve their thinking skills (Van Gelder, 2005). That is why, Miri, David, & Uri (2007) urged a paradigm shift in contemporary science education, by encouraging refrainment of implementing predominantly spread low-order thinking skills, and substituting them with higher-order thinking skills.

Chang, Chen, Lin and Sung (2008) stated that simulation activities were more beneficial for students with high-level reasoning skills. Regardless, Lewis & Smith (1993) urged that teaching of higher-order thinking skills should be implemented for all students, irrespective their learning abilities. Indeed, all adults at some point of their lives encounter situations where they need to use their problem-solving skills in order to tackle the tasks at hand. At that moment, it will not matter whether an individual was a gifted student or a student with learning difficulties in the past; what will matter is the individual’s ability to consider possible solutions to the problem. This example highlights the necessity of providing effective learning opportunities for all students in preparing to join the 21st century workforce. Similarly, the Miri, Davit and Uri (2007) study showed improvement in students’ critical thinking skills and boosted self-confidence through the practice of a vigorous curriculum and use of real-world problems.

Likewise, students’ critical thinking skills were expected to improve as a result of testing students on those skills (Barak & Dori, 2009). This expectation aligns well with Barak’s (2007) findings that constant testing of students on lower order thinking skills could lead to unwanted learning outcomes by reinforcing a superficial approach to learning. Therefore, a shift from the lower order thinking skills to the higher level thinking skills in students’ assessment is critical (Barak, 2007). Furthermore, in order to understand and address environmental issues in science, students not only need rich content knowledge but also a wide array of skills, such as higher-order thinking (Dresner, De Rivera, Fuccillo & Chang, 2014). Interestingly, several studies in the past revealed that some low achieving students showed resistance toward activities requiring higher order thinking skills (Halpern, 1998) and found these activities to be unfair and confusing (Williams, Oliver & Stockdale, 2004). Therefore, the 21st century skills must be explicitly taught; otherwise, students cannot develop them by themselves (Schlecher, 2012).

The instructional model used in the current study allows for both the teaching and the provision of practice of those skills. This new style of pedagogical instruction is imperative for students' future success, especially taking into account how current employers demand complex thinking skills over basic skills for the workforce (Levy & Murnane, 2005). Since critical thinking skills are so important for highly effective employees (Erwin & Sebrell, 2003), many colleges are predisposed to teach these skills in order to prepare their students for future, competitive work. Indeed, the memorization of formulas is not favored in colleges anymore, especially not in upper-level courses. This is because those low-level tasks preclude students' problem solving skills (Chiel, McManus, & Shaw, 2010). Jagger (2013) stated that students should be able to achieve Bloom's taxonomy's higher cognitive levels by senior year of college. Even though colleges' and researchers' keenness for teaching critical thinking skills to college students is laudable, this position is not without controversy: it lowers expectations on students' achievement during their high school years. Indeed, teachers and instructors should be cognizant of the importance of these skills while the students are in school. The expectation for all students should be to achieve mastery of higher order thinking skills before beginning college. Building on learners' pre-established knowledge requires critical thinking, which enables the learner to take on perspectives from different angles, respect diverse ideas and tailor evidence-based scientific arguments (MacKnight, 2000).

Technology and science learning. The integration of technology into curriculum is another way to foster students' critical thinking. It helps students take on analysis and synthesis while working on authentic tasks (Harris, 2005). Educational software similar to River City was used in the past. For example, students learned life science by utilizing teaching agents, where learning took place via teaching (Biswas, Schwartz, Bransford, & TAG-V, 2005) and through guided discovery video games (Schwartz, Blair, Biswas, Leelawong, & Davis, 2007). In contrast with schoolwork, which may be deemed as boring, games are fun and pleasant activities for students who may otherwise be unmotivated (Burn, Buckingham, Parry, & Powell, 2010). Therefore, learning with games is a useful pedagogical methodology for encouraging otherwise apathetic students to learn. Furthermore, Mayo (2007) speculated that video games have a big potential to deepen and broaden students' scientific knowledge; the games provide students with opportunities to work on real-world phenomena, which require reasoning rather than memorized facts in order to tackle challenging problems.

Furthermore, Mayo argued that game-based learning might help the U.S. produce more scientists and engineers. It is worth noting that studies done in the past evaluated the impact of video games on brain activity and revealed that video games stimulate the production of a neurotransmitter called dopamine, which is an important chemical used by the human memory system (Koepp, Gunn, Lawrence, Cunningham, Dagher, Jones, Brooks, Bench, & Grasby, 1998). As we already know from research, dopamine signaling stimulates learning (Byrne, Patrick, & Worthy, 2016).

Swiftly growing literature regarding technology integration in science learning indicates increased robust understanding of science among high school students. The inclusion of new innovative technologies within learning improved education in general and student achievement in particular (Johnson, Levine, Smith & Haywood, 2010). Also, integration of technology in virtual laboratory settings was considered an advantageous tool for science learning (Swan & O'Donnell, 2009). Adapting technology in the classrooms along with fostering student curiosity, imagination, collaboration, complex thinking, problem solving and information analysis, teaches students 21st century skills that are essential in the current globalized world (Saavedra & Opfer, 2012). The Galapagos Finches software integrated in biology lessons (Tabak & Reiser, 2008) served as a pragmatic tool for creating a computer-based inquiry environment, where students demonstrated progress in their content knowledge regardless of their previous academic success. Likewise, Liu and Hmelo-Silver (2009) utilized hypermedia technology with their students in order to study the structure-behavior-function (SBF) conceptual representation, which improved students' understanding of complex systems and, therefore, assisted in organization of knowledge (Pea, 1993). Scheuer, McLaren, Weinberger and Niebuhr, (2014) designed computer-based collaboration scripts, which utilized hybrid software to support both the argument diagraming and scripted argumentative discourse. As a result of these scripts, student discussions were enriched by a significant number of elaborative moves. Furthermore, researchers acknowledged that implementing technology-enhanced instruction has a positive impact on English Language Learners' (ELL) science learning (Ryoo, 2015). In particular, Ryoo's (2015) study showed that both ELL and non-ELL students were able to develop a richer understanding of complex scientific phenomena when they learned through web-based, interactive instruction, rather than merely using textbooks. Several authors called for the use of special digital tools and learning environments that enhance the effectiveness of online collaborative learning. These tools and environments included a group coordination tool (Kwon, Hong, & Laffey, 2013), personalized virtual learning spaces called PVLSs (Kallkvist, Gomez, Andersson, & Lush, 2009), a digital playful learning environment called PLE (de Koning-Veenstra, Steenbeek, van Dijk, & van Geert, 2014), and an online peer questioning through Wiki (Cho, Lee, & Jonassen, 2011). The effectiveness of these approaches in online collaborative learning suggests that in order to make online questioning more interactive and engaging, it may be useful to design a digital tool and/or environment to facilitate online questioning.

Students' questioning techniques. Question asking and answering are considered key elements in theories of learning, cognition and education (Graesser & Person, 1994). When students encounter situations where more than one possible choice is available, or when they are puzzled about scientific phenomena and want to find out more, asking questions becomes a natural habit of the mind. This helps them recognize the strengths and weaknesses of their ideas and, therefore, helps them evaluate the evidence supporting or refuting their hypothesis (Chin & Osborne, 2010). MacKnight (2000) advocated for the engagement of students in online discussions and teaching critical thinking skills through Socratic questioning prompts. The River City project allowed students to participate in a mysterious problem solving activity while working on enhancing their questioning skills through the Goggle Docs web-based application.

Question asking is important for the successful processing of critical information and helps in achieving better comprehension (Costa, Caldeira, Gallastegui, & Otero, 2000). Moreover, good question asking skills are required for development of scientific thinking, and are an important part of scientific inquiry and science practices (Brill & Yarden, 2003; National Research Council, 1996; Next Generation Science Standards, 2013). Higher order questions necessitate high cognitive demand, deep thought and reasoning (Bloom, Engelhart, Furst, Hill, & Krathwohl, 1956), whereas questions requiring recall or mere retrieval of information are considered low level questions (Foote, 1998). Thus, higher order questioning is a powerful tool, which enables students to think critically, in contrast with lower order questioning, which relies on memorized facts (Wimer, Ridenour, Thomas, & Place, 2001). Likewise, the quality of a question can be determined based on the answer it entails (Yarden, Brill, & Falk, 2001).

King (1992) argued that questioning is a powerful pedagogical methodology: whenever students used question stems and constructed higher level thinking questions, they outperformed students who did not use questioning techniques. By posing thought-provoking and probing questions, teachers foster higher-order thinking skills among students (Saavedra & Opfer, 2012; Schwartz & Fischer, 2006). Cuccio-Schirripa and Steiner (2000) considered questioning to be a key integrated component of critical thinking, problem solving, and creative thinking. Furthermore, these authors argued that students who received training on researchable questions were able to generate high quality questions, unlike their peers in the control group who did not. High school students' training in high quality questioning techniques is necessary, since the study conducted by Choi, Land, & Turgeon (2005) evaluated circumstances where college students were not able to generate high quality questions that required deep-understanding and reasoning, and formulated low quality questions instead. Furthermore, high quality, student-crafted questions promote interactive learning during peer tutoring through constructive, deep, thought-provoking questioning and elaborated explanations (Cho et al., 2011; Ismail & Alexander, 2005). The above statements align well with the idea that when learners experience a state of cognitive disequilibrium (Graesser, Lu, Olde, Cooper-Pye & Whitten, 2005), it prompts them to ask questions (Bradley, Thom, Hayes, & Hay, 2008). Furthermore, Bradley and colleagues argued that the type of questions asked by undergraduate college students affected the quality and quantity of their online submissions. It should be noted that surprisingly recent studies revealed that college undergraduates placed in low-level embedded sentence-generating tasks were able to construct more task-relevant knowledge and more accurate sentences, than their counterparts who were placed in higher-level queries (Cuevas & Fiore, 2014). This finding contradicts the claims mentioned above.

Ford's (2008) study draws a distinction between how scientific knowledge is constructed by scientists and the reasoning involved in learning content with understanding. The model developed in that study emphasized the importance of students knowing how to hold claims accountable to scientific practice and reasoning in contrast with dominant constructivist ideas that emphasize knowledge construction by students acting like scientists. The construction of knowledge is an important aspect of constructivist theories (Palincsar, 1998). Students construct knowledge and, therefore, practice a constructivist style of learning when they generate their own questions (Watts, Gould, & Alsop, 1997). The small portion of this study builds on prior research that demonstrated that students are capable of generating thoughtful questions in a science context; moreover, question asking leads to critical reasoning (Chin & Osborne, 2008).

One of the focus areas of this study is the construction of student-generated questions and the quality of these questions. Chin and Osborne (2008) stated that student generated questions are important in both science learning and science teaching. Dillon (1988) argued that students do not spontaneously generate high level thinking questions. Quite contrarily, students formulate high quality questions when they receive special training or work in an encouraging environment. King (1989, 1990, 1991) suggested an instructional strategy where students craft their own questions based on the question stems which she adapted from Ryan (1971). Moreover, in another study, King and Rosenshine (1993) found that, during science discussions, students who used more elaborated question stems (What would happen if…?) outperformed students who used less elaborated question stems (Why…?).

Furthermore, although question generation is a useful cognitive strategy, as it increases learner’s comprehension (King, 1992a; Rosenshine, Meister, & Chapman, 1996), Lacasa, Martinez, & del Castillo, 2011 observed the benefits of student questioning in different dimensions and argued that question generation alone during oral dialog was not as effective a pedagogical strategy as students’ reliance and reflection on the written language.

One of the pedagogical strategies that can foster student question asking is question-driven problem-based learning (Q-PBL), which can be successfully implemented in science classrooms (Chin & Chia, 2004). Gross (2001) categorized students’ questions into two different types: a) self-generated questions spawn by intrinsic motivation, and b) extrinsically motivated questions posed by their teachers, textbooks or peers. Chin and Kayalvizhi (2005) results indicated that students preferred to use their own self-generated questions rather than textbook questions. Likewise, research found that students' self-generated questions contributed to meaningful science learning (Chin, Brown, & Bruce, 2002; Chin & Osborne, 2008). A similar approach was used by Gelmini-Hornsby, Ainsworth, & O'Malley (2011), whose study on the Guided Reciprocal Peer Questioning (GRPG) revealed that question prompts enhanced computer-supported collaborative learning and encouraged students to construct more elaborated explanations. Similarly, Zoller (1994) developed and implemented a pedagogical strategy called ESAQ (Examination where Students Ask Questions) for undergraduate students, which led to increased student motivation and participation.

Also, research revealed that student generated questions were helpful in guiding students' thinking and writing during their writing tasks, including answering essay questions and writing project reports. Science Writing Heuristics (SWH) was used to facilitate students' write-to-learn activities (Wallace, Hand & Prain, 2004). According to Otfinowski & Silva-Opps (2015), eliciting questions in science allows students to question their pre-existing scientific knowledge, which consequently increases students' confidence in scientific writing. It also provides students with a realistic idea of the process of conceptual change in science (Vosniadou, 2013), as well as a shift in students' science identities. It is hypothesized that the usage of higher-order thinking questioning should enhance students' science identities. Costa et al. (2000) analyzed the quality and the quantity of questions raised by students who were evaluating a scientific text and found that students were capable of asking many questions if they were given opportunities to do. Students' self-generated questions were used as indication of their scientific interest in order to develop a holistic approach to science learning (Baram-

Tsabari, Sethi, Bry, & Yarden, 2006). Past research showed that students asked few questions in classroom settings and that these questions were mostly low cognitive level questions (Dillon, 1988). Unless students were trained to use high quality questions, only a few of them asked questions with high cognitive demand (Carr, 1998).

Urban Education. Urban education is riddled with problems like inadequate funding, poor teaching quality, unequal socio-economic conditions, and dysfunctional neighborhoods with high crime rates. Modern school reforms try to address these problems by taking either comprehensive, nation-wide approaches or by implementing community-specific strategies. For example, the Comprehensive School Reform (CSR) refers to the agenda of implementing comprehensive school reforms, which include the development of effective instructional practices, the improving of curriculum and assessment, and the supporting of the community partnership programs which strengthen parent and community involvement in education (Slavin, 2008). The School Development Program (SDP), one of the earliest school intervention programs designed by an African-American Yale psychiatrist, was focused on improving the test scores, behavior, and attendance of poor and/or socially marginalized students. This program relied on a connected community and parent population to help toward improvement of students' behavior and motivation (Comer, Haynes, Joyner, & Ben-Avie, 1996).

Currently, urban education reform attracts more attention from policy makers, educators and researchers than any other type of school reform. Since the 1980s, almost every large school district adopted some form of market-driven reform (Lipman, 2004). Anyon (2008) argues, school reform on its own may not be a full solution for problems taking place in urban education, unless economic reform is integrated as well.

Furthermore, as Anyon (1980) suggested, student work in working-class classrooms tends to be quite different from student work in middle-class or affluent classrooms. Indeed, in working-class schools, education is often simplistic and mechanical with students spending their time on taking notes and memorizing facts from their books. Not enough time is dedicated to experiments and analytical examination of learning material.

In addition, the classroom discourse in these schools tends to be primitive without good questioning and explanations. Rubin's (2007) study explored the deficiencies of classroom discourse in urban schools. The results of this study indicated that problems with discourse and student interactions led to a learning environment where most students were predestined to fail. The study showed the urgent need for better discourse and interactions in order to improve the learning environment and help students succeed. Zohar and Dori (2003) found out that many teachers believed that higher order thinking activities were not suitable for low achieving students. As a result, they altered their instructions toward lower order activities while teaching low achieving students. This is regrettable, as the research confirmed that activities requiring higher order thinking skills can be beneficial for low achieving students: "The compelling empirical evidence shows that low-achieving students and higher order thinking are not mutually exclusive" (Zohar & Dori, 2003, p. 177).

As Andersen and Ward (2014) argued, minority students "may internalize the idea that they cannot perform science or may feel that they must lose their racial identity to be assimilated into the culture of science." Therefore, the use of culturally cognizant science lessons is important, since research indicated that the integration of culturally responsive teaching approaches into science lessons positively affected minority students' motivation (Hines, 2003).

Science Identities. Development of a self (identity) starts in the early developmental stages and carries on to adolescent years (Moje, Tucker-Raymond, Varelas, Pappas, Korzh, & Wentland, 2007). As Moje argued, young people struggle to figure out who they are and what they want to become in the process of "achieving" an identity. Eli and his colleagues (Moje et al., 2007) disagree further, that it is not about "achieving" or "having" scientist identities, but rather about "awareness" of scientific activities, where students have the chance to tell stories and reflect on them. However, again addressing the question of at which age science identities develop, research says that it is not conceived in high school, but rather in primary school (Moje et al., 2007). On the contrary, Muuss (1996) claimed that late adolescent years are very important years for individuals' identity formation. If students do not view themselves as capable of engaging in scientific inquiry, then they will not engage in it. This will result in their exclusion from the technologically advanced job market workforce (Moje et al., 2007). Identity can be considered an expression of self, and "identities shape and are shaped by practices…" (Moje et al., 2007, p. 595). Identities are constantly shifting and evolving, although research in the past stated that identities are too stable (Burke, 1980).

Is there a difference between science and scientist identities? If one has a science identity, isn't it automatically considered a scientist identity? For instance, Eli, Maria and Chris (Moje et al., 2007) distinguish two different types of scientist identities: actual and acting. Furthermore, the deficit of sufficient explanation of scientific question followed by "I do not know" is explained by lack of possession and construction of designated scientist identities (Moje et al., 2007). Moje questioned whether the construction of science identities produced “better scientists” and helped students learn science. Additionally, Moje asked “if one must really take on a scientific identity in order to engage with scientific activities, or does one simply need to understand the epistemological stances of science, the attitudes, beliefs, values and ultimately, the discourses privileged in scientific inquiry in order to engage with science?” (Moje et al., 2007, p. 596). These are valid questions that need to be answered. Although there were several studies conducted regarding science identities and various researchers contributed ideas toward the notation, definition, implementation and integration of science identities, it seems that there still are gaps that need to be filled by conducting more research in diverse settings.

Carlone and Johnson (2007) designed a model of science identity that evaluates students’ science identities across three overlapping dimensions: competence, performance and recognition. This model claims that individuals with strong science identities rate themselves highly and are rated highly by others, in all three dimensions. However, the model allows for variations in the individual’s identity composition across these dimensions. According to Carlone and Johnson (2007), recognizing oneself and being recognized by others as a “science person” give the individual a strong science identity. For this reason, Andersen and Ward (2014) placed strong emphasis on the improvement of students’ own perceptions regarding their science identities. Likewise, Kane (2012) found there to be a critical connection between science content learning and the construction of science identities.

Viewing identities, identity construction and learning together provide a new, powerful perspective for assessing and understanding learning as a process (Varelas, 2012). Various researchers considered the concept of science identity to be an analytic lens for studying science learning as a socialization process, rather than as a buildup of knowledge (Gee, 2000; Carlone & Johnson, 2007). The identity lens opens up new avenues for researchers to understand teaching and science learning environments, by allowing science educators to understand why or why not students find learning science worthy (Carlone & Johnson, 2007).

The last but not least prodigious factor about the construct of science identities is the forms and sources of data analysis. In order to achieve to sharper perception of valid science identities' analysis, interviews conducted with students, shadowing and observing students in various situations shed lights on accurate data analysis and data interpretation. This model provides an important bridge between theory and research (Moje et al., 2007).

Method. The current study used qualitative data collection methods that aimed at understanding urban youth's science identities and motivation in science. Carlone and Johnson (2007) science identities framework was utilized for analyzing students' SI. Would the River City project provide any empirical evidence about students' SI? Are web-based science activities the pathways to development of stronger identities? The study aimed at answering the following research questions (Q) and sub-questions (SQ):

Q1. How do web-based science activities influence students' science identities?

SQ1. What conclusions regarding students' science identities can be drawn from students' letters to the mayor of River City?

SQ2. Based on the data of the SIEVEA instrument, how can students' science identities and motivation in science be explained?

SQ3. How can the data from the developed questionnaire be used in evaluating high school students' science identities and motivation?

Q2. How do urban students' science identities change throughout inquiry-based online collaborations?

Q3. Do question stems always help students generate thoughtful and meaningful questions?

In order to investigate these questions, several pieces of data were collected using the following three qualitative research tools:

1. SIQ (Science Identities Questionnaire) that served as a substitute for paper and pencil interviewing (PAPI),
2. SIEVEA survey instrument used for analyzing students' science identities and motivation in science,
3. Students' letters to the River City Mayor written at the beginning and toward the end of the project.

Students took the SIQ questionnaires in their science classrooms twice: at the beginning and at completion of the River City project. The questionnaire contained nine open-ended questions and was presented in interview format.

The SIEVEA survey was administered in paper and pencil. Both the pre- and post- tests were administered. This survey was based on a 5-point Likert scale and used the following answer choices: (1) Strongly Agree, (2) Agree, (3) Neither agree nor disagree (4) Disagree, and (5) Strongly Disagree. The first section of the survey measured students' science identities and motivation in science, the second section measured students' values of science and the third section measured students' environmental attitudes (see Figure 4.2). In this study, only the results of the "science identities and motivation in science" part of the instrument were analyzed. Data collected was used for answering the first and second research questions related to changes in students' science identities and motivation in science. The other two sectors, the values of science and environmental attitudes were not in the researcher's interest, so they were excluded during data analysis.

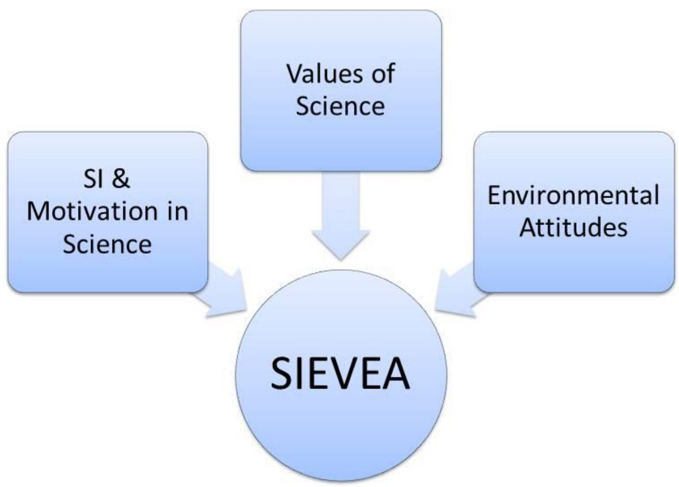


Figure 4.2. SIEVEA Survey Instrument

After taking a tour around River City, interviewing various residents and making observations, students were asked to write a letter to the town's mayor. In their letters, students were asked to inform the mayor about the situation in River City and to propose a solution to the problem. Students were asked to write another letter to the mayor of River City after finishing the project. These letters were also used for evaluating students' science identities.

Students' questions from the Google Docs and the mayor's letters from the River City MUVE were used to explore the research questions related to the construction of meaningful higher-order thinking questions and students' science identities and motivation in science.

Participants and settings. Initially, 33 students agreed to participate in this project. All participants had school-provided MacBook Air laptops, but not all of them had Internet access at home. Additionally, while the River City software ran smoothly on the Windows platform, it had some errors on the Mac platform. Therefore, the student pool had to be reduced to include only those students who had Internet access at home and Windows PC. This cut down the sample size to 20. Then, some students voluntarily dropped off the project after realizing the time and effort needed to commit to the project. The final sample consisted of eight students. As an incentive, these students were awarded extra credit by their science teacher and volunteering hours through the university. The River City program sent daily emails to the researcher, showing all the work and time each student spent on the project. The volunteering hours were calculated based on the data in those emails.

The eight study participants were four boys and four girls. All were ninth grade urban high school students enrolled in the Biology course. The study was not conducted in a classroom setting, but rather in an online-collaborative environment, where students were required to log in from home and participate in the study during after school hours. Students spent about five hours per week on the project, not including additional postings and discussions that took place in the Google Docs. The total duration of the project was two weeks.

Research design. This qualitative study used a grounded approach (Glasser & Strauss, 1967) for comparative analysis of groups' qualitative data and for identifying any emerging patterns. The purpose of this grounded theory study was to investigate and analyze urban youth's science identities.

The grounded theory was chosen as the method of the qualitative research for this study due to data collection, analysis and a subsequent formation of the theoretical model (Glaser & Strauss, 1967). Creswell (2009, p.229) argued that grounded theory is a "qualitative strategy of inquiry in which the researcher derives a general, abstract theory of process, action, or interaction grounded in the views of participants in a study." Data, used for this grounded theory study, included students' answers to the questions of the SIEVEA survey, open-ended, and closed questions that were associated with the SIQ and students' letters to the Mayor of the River City. The study participants were from a high school located in the East Coast, studying biology.

During the study, students worked on real-world problems in an online collaborative learning environment, which simulated genuine scientific activities. The River City is a multi-user virtual environment (MUVE) for teaching students scientific inquiry and 21st century skills in their science classes. It is based on the National Science Standards (National Research Council, 1996) and features environmental and epidemiological topics. The human impact on environmental health is also in the center of Next Generation Science Standards (NRC, 2013).

In this project, students acted as River City visitors. As visitors to River City, they traveled back in time and brought their 21st century knowledge and technology to address 19th century problems. River City is a town plagued with health and ecological problems, and students worked together in investigating the root causes of these problems. They collected data, formed hypotheses, developed controlled experiments to test their hypotheses, and made recommendations based on their findings.

The River City project engaged students in scientific inquiry through meaningful and relevant real-world problem solving tasks. This study used the constructivist pedagogy; through the inquiry process, students acquired skills that scientists use for making inferences leading to valid claims (Colucci-Gray, Camino, Barbiero & Gray, 2006). The project utilized an educational video game format (Saavedra & Opfer, 2012) and a 3D immersive virtual environment (Ketelhut & Nelson, 2010) as teaching tools.

The project worked as enrichment, after-school program. Since it was like a video game, there were no exact directions what to do next and how to find that particular piece of information. In lieu of the directions, the researcher modified and simplified the guide provided by the River City to help students navigate throughout the town. The guide was posted on Google Docs. Since in the past this project was carried out mostly with middle-school students and its design and national standards were aligned with up to 12th grade curriculum, it was a good idea to try it with high-school freshmen while providing less support and scaffolding.

Procedures. The study was comprised of multiple steps/phases (see Figure 4.3). During Day 1 of the study, the researcher introduced the project to all participating students. The twenty minute presentation covered the objectives of the study, its benefits for students, its duration and detailed schedule.

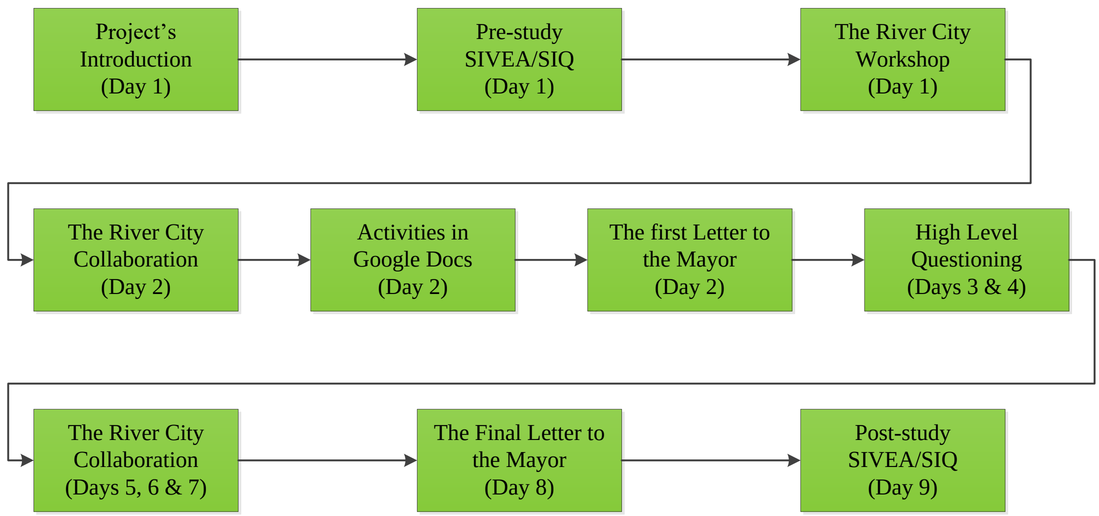


Figure 4.3. The Study's Steps/Phases in a Chronological Order

Following the introduction, all participating students completed the SIEVEA survey and the SIQ questionnaire (see Appendix B). The last part of Day 1, about 20 minutes, was a workshop during which the students learned how to connect to the River City MUVE and how to use tools within this environment.

During the following two weeks, students worked on the River City project. During the entire project, students were engaged in activities that required problem-solving skills, conducted research, tested hypothesis and conducted virtual experiments. The study was not conducted in classroom settings, but rather in an online-collaborative environment, where students were required to log in from home and participate in the study during the after school hours. Students worked on the project using online collaborative tools and their discussions were captured in their online postings. Students collaborated via postings in the River City MUVE and Google Docs (Google Docs is a web-based word-processor, which allows students to create and edit documents and engage in collaborative activities; see Appendix F). Students were asked to log into Google Docs twice. Every morning the researcher received email reports regarding students' notebook entries and embedded assessments. The daily email was very specific; it included the student name, the time that student was in River City and his/her actual activity (the visited places, the names of people who were engaged in communication, what questions were asked or answered, and so on). A sample email with the changed avatar name can be found in Appendix G.

The online setting of the project allowed the researcher to introduce students to the questioning skills, which was based on Chin's (2006) questioning framework, including a) modeling higher-order thinking questioning asking, b) providing question stems, c) providing a question taxonomy (Bloom's), and d) asking students to practice these skills by posing and answering these questions on Google Docs.

The materials, including the tutorial in the form of the PowerPoint, which explained the difference between higher-order thinking and lower-level questioning, as well their impact on science learning, were posted on the Google Docs. Likewise, Alison King's question stems, the practice problems and the New York Times article that students were asked to use for practicing good quality questions were also posted on the Google Docs. The author has conducted two preliminary studies regarding student questioning techniques prior to conducting the present study. These studies indicated that question-stems help students in generating good quality questions. For the current study, the researcher's aim was toward understanding the meaningfulness of the constructed higher-order thinking questions. Students were trained on the questioning techniques and had a chance to generate their questions based on the New York Times article and River City/school science topics. The examples and the utilization of higher-order thinking questions in science learning context were also posted on the Google Docs. Students were given question prompts for practice purposes. This was similar to the previous preliminary study (Aghekyan, 2015) where students were asked to recognize, differentiate and construct higher-order thinking questions. In addition, current research involved online collaborative tools, which was a suitable pedagogical strategy for the study's subjects, since social learning is minority students' preferred learning method (Heilbronner, 2011).

In the last day of the project, students wrote their post-letters directed to the mayor of River City. Additionally, students completed the SIEVEA survey and the SIQ. The SIEVEA post-test had the exact same questions as the pre-test, but in a different order. The SIQ included some questions that were asked during the pre-test, but most of the questions were relevant to River City. The pre- and post-test data of the SIEVEA and SIQ along with the mayor's pre- and post-letters were compared for understanding how the urban youth's SI and motivation in science evolved during the study.

Data analysis. *Students' letters to the Mayor*. Students were asked to write a letter to the mayor of River City regarding the unfolding crisis in the city. Students were encouraged to provide an update as well as make recommendations regarding the improvement of River City's environmental health. Furthermore, students were asked to write the letter twice; one letter at the beginning and another at the end of the project. It was conceivable that students would gain more knowledge toward the end of the project. However, it was not the length of the letter, nor the details of the described situation that was used for evaluating the letters to the mayor.

Before writing the first letter to the mayor, students had a chance to explore River City, interview the residents of the town, and navigate through the tenements and wealthy houses, university and the hospital while making meticulous observations and inferences. River City modeled real-life problems and engaged students in decision-making activities, which assisted them with writing mayor's letters. After initial interaction with the River City, students wrote a letter to the mayor of the city regarding the ongoing issues with the River City. After the students progressed through four different seasons (October 1878, January 1879, April 1879, and July 1879) and completed all activities of the project, they wrote the second, final letter to the mayor of River City in which they explained their findings and made recommendations for improving River City's environmental health.

To keep the text concise, the pre-letter and post-letter terminology was used when evaluating the letters and discussing their scores. The Carlone and Johnson (2007) SI framework distinguished three different aspects of SI: performance, competence and recognition. For analyzing the letters, the performance and competence aspects of SI were considered, whereas the recognition part was excluded, because it was highly unlikely that students would mention anything in the mayor's letters about their science recognitions. The competence category consisted of the following subcategories: knowledge of the science content, understanding scientific phenomena, motivated to understand the world scientifically and understanding the scientific method. The performance category was divided into the following subcategories: usage of scientific tools, fluency with the scientific talk, acting as scientist and confident interactions in formal and informal scientific settings. For the complete matrix called **S**cience **I**dentities: **C**ompetence **a**nd **P**erformance (SICAP), see Appendix A.

While analyzing the competence and performance aspects of SI, the all-or-none approach was used: for each specific subcategory of science identity, it was assumed that the subcategory was either present or not present in the letter (see Appendix C for students' letters to the mayor). Therefore, each subcategory of competence and performance was assigned a score of zero or one. The zero score indicated an absence of a particular subcategory, whereas the score of one represented a presence of that subcategory. For example, if the student's letter showed fluency in scientific talk, the letter received a score of one in that subcategory. Otherwise, it received a zero. It should be noted that the relative strengths of competence and performance subcategories were not analyzed. For instance, the researcher was not concerned with how strong or weak the student's scientific talk was.

In order to avoid any possible bias, students' written letters were analyzed by two evaluators who assigned numeric scores to each letter in each subcategory of SI. Then, the letter's average score was calculated and used as the letter's overall score. The final scores of letters were used for comparing the pre- and post- letters. The following simple formula was used for calculating each letter's total SI scores: SICAP = competence + performance. For each letter, eight was the maximum possible score, whereas zero was the lowest.

Student generated questions. Kuhn and Reiser (2006) stressed the importance of creating an environment and need for engaging students in scientific argumentation. Equally, there is an urgent need for creating questioning opportunities for students, since students learn more by asking effective questions rather than by memorizing facts (Toledo, 2015). Research claims that both question asking (King, 1994) and question answering (Roscoe & Chi, 2008) are beneficial for the learning process. Although questioning techniques used by inquiry teachers tend to provoke students' deep thinking (Hmelo-Silver & Barrows, 2006), this study focused on student questioning, more specifically, the meaningfulness of their constructed questions.

In this study, students were introduced to question stems and question classification into lower and higher level questions. Bloom's taxonomy provides a suitable framework for this type of classification (Bloom, Engelhart, Furst, Hill, & Krathwohl, 1956). Within Bloom's taxonomy, student-constructed questions are grouped into two groups: lower level and higher level-questioning. These groups themselves are divided into subcategories, based on six different expertise levels defined within Bloom's taxonomy. They are manifestations of the measurable outcomes of student learning, which show their mastery of a particular domain of knowledge. Accordingly, lower level thinking questions are cataloged into knowledge, comprehension, and application subcategories, whereas higher level thinking questions fall into the analysis, synthesis, and evaluation areas (Bloom, Engelhart, Furst, Hill, & Krathwohl, 1956).

This study used the revised version of Bloom's taxonomy, created by Bloom's former graduate student, Lorin Anderson and his colleagues, in order to reflect on the new skills required for 21st century students (see Figure 4.4 for more details). The major difference between Bloom's original and revised taxonomy is that the original taxonomy has a single dimension, whereas the revised taxonomy has dual dimensions: learning and cognition (Anderson, Krathwohl, & Bloom, 2001). In order to code and analyze the questions, "Questions to provoke critical thinking" table, created by the Harriet W. Sheridan Center for Teaching and Learning of Brown University, was used. The Sheridan Center's table was based on the thinking skills and example question stems from Alison King's (1995) "Inquiry Minds Really want to Know: Using Questioning to Teach Critical Thinking."

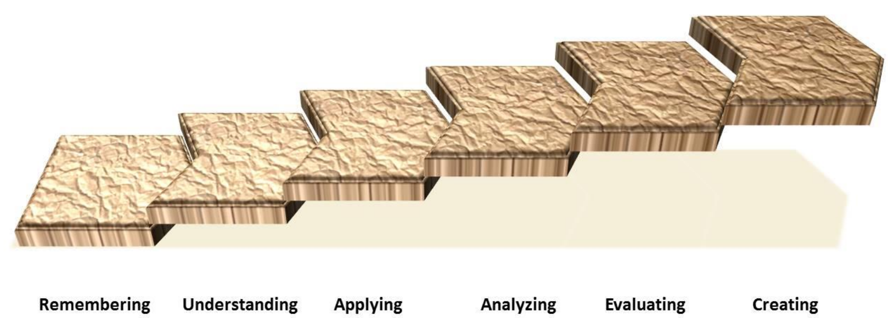


Figure 4.4. Bloom's Revised Taxonomy

The study allocated two days for training of students on questioning techniques. During these two days, students did not work on the River City project, but rather utilized Google Docs application that served as a teaching and learning tool for constructing questions. The researcher posted a Power Point presentation, which explained the importance of constructing and utilizing higher-order thinking questions, as well as illustrated the generation of these types of questions. There was also a Word document posted on Google Docs, which allowed students to practice their learned skills. Likewise, Alison King's (1994) question stems were posted as a guide for crafting high quality questions. Students were asked to review the posted question stems and to apply them to generating higher-order thinking questions in a science context. The researcher's goal was to find out whether students were able to construct meaningful, high quality questions when they were given the question stems. For that purpose, all student-generated questions were used for detailed qualitative analysis through the lens of meaningful question formulation.

Results. ***Students' letters to the Mayor.*** The evaluation of the sixteen letters to the mayor produced the following cumulative scores: reviewer 1 (pre-test = 40 and post-test = 53) and reviewer 2 (pre-test = 30 and post-test = 45). The mean score of the pre-test letters was 4.375, whereas the mean score of the post-test letters was 6.125. The data showed that there was a significant change in scores, with post-letters showing an increase in students' SI in comparison with their pre-letters. Table 4.1 summarized pre- and post-test scores of students' letters to the Mayor of River City.

Table 4.1
Scores of Students' Letters to the Mayor of River City

Student	Pre-Test	Post-Test	Difference	Relative Change
Lea (S1)	4	6.5	2.5	.625
Shanaia (S2)	3	5.5	2.5	.500
Samad (S3)	4.5	6	1.5	.429
Tom (S4)	5	7.5	2.5	.833
Tanashia (S5)	4	5.5	1.5	.375
Jose (S6)	6.5	7.5	1	.667
Juan (S7)	5	6.5	1.5	.500
Jessica (S8)	3	4	1	.200

As the data in Table 4.1 showed, all students experienced an increase in their SICAP scores. Both the competence and performance categories of SI were present in students' letters at the beginning of the project, although not every subcategory of the SICAP matrix was present. Lea (S1), Shanaia (S2) and Tom (S4) experienced the biggest change: 2.5 points. Samad (S3), Tanashia (S5) and Juan (S7) had the second largest change of 1.5 points. Lastly, Jose's (S6) and Jessica's (S8) scores increased by 1.

Even though it appeared that Lea, Shanaia and Tom had the largest increase, it was important to put this into a proper context, by looking at the change relative to students' initial abilities. Indeed, since Shanaia's pre-test score was 3 out of a maximum of 8, it was easier for her to increase her score by 2.5 than for Tom, whose initial score was 5 out of 8. Shanaia had a room of 5 points for growth (maximum 8 – initial 3 = 5). However, Tom's had only 8 – 5 = 3 points for possible increase. Therefore, a new measure, called "Relative Change," was introduced to account for differences in students' initial (pre-test) scores. The formula for calculating this new metric was rather simple: (post-test score – pre-test score)/(maximum score (8) – pre-test score). For example, for Shanaia this metric equaled to (5.5 – 3)/(8 – 3) = .5.

According to the values of relative change, Tom had the biggest increase in his scores: .833. His scored increased from pre-test 5 to post-test 7.5. The next largest value was Jose's score: .667. This was due to his pre-test score of 6.5 becoming 7.5. Even though Jose's absolute score change was modest, only 1 point, he only had room of 8 – 6.5 = 1.5 for an increase. This means that Jose was able to add 1 score out of the possible 1.5, which seemed to be a significant change. Next, Lea was able to increase her score from pre-test 4 to post-test 6.5, resulting in .625 of relative change.

Jessica had the smallest increase in her scores. She scored a meager 3 on her pre-test and 4 on her post-test. Tanashia did not impress either by scoring 4 and 5.5 on her pre- and post- letters. The remaining three students, Juan, Shanaia and Samad, had score increases in the middle range of all scores.

The analysis of students' letters to the Mayor of River City indicated that during the project students' science identities changed to some extent. However, it should be noted that the significance of this change was not clear due to the small sample size, the short duration of the project (2 weeks) and measurement errors. Moje et al. (2007) study has a particular relevance here, since it also stated that students' science identities were not stagnant, but rather could change.

Student generated questions. In the middle of the project, students were introduced to questioning techniques. There were several materials posted on Google Docs. For instance, a posted PowerPoint explained the difference between the higher-order and lower-order thinking questioning and highlighted the advantages of utilizing good quality questions in discussions. There were four other Word documents posted, including a document named "Problem of the day," New York Times article about the increased level of carbon dioxide, a handout which allowed students practice formulating questions, and a document with question-stems designed by Brown University using Alison King's question stems.

The previous research (Aghekyan, 2015) indicated that students were capable of constructing good quality questions when they were provided with question stems. However, Aghekyan (2015) noticed that students were not always able to generate sensible questions. Specifically, after students used the question stems, they were able to formulate questions, but some of these questions did not make sense. In this study, the researcher wanted to introduce students with the questioning techniques and evaluate the meaningfulness of student-generated questions. Therefore, the research question related to student-generated questioning was “Do question stems always help students generate thoughtful and meaningful questions?”

Before doing any analysis, all student-crafted questions were organized in a separate file. The students’ names were not tracked because the research did not look into linking student-generated questions posted on Google Docs with their performance in the River City project. Student-generated questions can be seen in Appendix E. The following coding schema was implemented: meaningful questions were coded as (M), whereas for not-meaningful questions (NM) code was used.

Out of twenty-seven student-generated questions, only three questions (#5, #7, and #13) were coded as non-meaningful. The remaining questions made sense and were both low-level and high-level questions. This confirmed the previous findings that, when question-stems were provided, students were capable of constructing meaningful, good quality questions.

SIEVEA survey – science identities and motivation in science. Students took the SIEVEA survey twice: as a pre-test and post-test. The SIEVEA instrument was developed and validated by conducting exploratory (EFA) and confirmatory (CFA) factor analyses and Rasch analysis. The pre-test was previously administered to 3,454 high school student participants in New Jersey, Pennsylvania and Connecticut. The post-test survey included the exact same pre-test questions but in a different order. The initial survey was designed to measure four constructs: science identities, expectations of success, values of science and environmental attitudes. After conducting factor analysis, it became clear that the survey consisted of three constructs: science identities and motivation in science, values of science and environmental attitudes. Although students completed the entire survey, only seven items related to science identities and motivation in science were used for the following analysis. The remaining survey items, regarding students' environmental attitudes and values of science, were not included in the analysis because researching these constructs was out of scope of this study.

During the qualitative data analysis, each pre-test item was compared to its corresponding post-test item. This methodology was feasible, since the items of the pre- and post- tests were the same; only their order was different. For example, the third item in the pre-test survey instrument was matched up with the twelfth item in the post-survey; the twelfth item of the pre-test was matched up with the post-test survey's item eleven and so on. This procedure was done for every student for all items concerning students' science identities and motivation in science. Each student answer was assigned a rating of one through five using the 5-point Likert scale. Then, these ratings were used to compare pre-test survey results to the post-test results. Based on the comparison results, students' answers were divided into three groups: positive changes, negative changes and no changes. Table 4.2 summarizes the SIEVEA's pre- and post-test results. For each item, the changes columns indicate the number of students whose pre- and post-test results on that particular item showed a positive, negative or no change.

Table 4.2
Comparisons Between SIEVEA Pre- and Post-Test Results

Item	Positive Changes	Negative Changes	No Change
Learning science in school will help me to succeed later in life.	1	1	6
I am confident I can master the skills taught in my science class.	0	4	4
I consider science topics very interesting and engaging.	1	5	2
When it comes to learning science, I think of myself as a science person.	4	3	1
My peers and teachers think that I am knowledgeable in science.	1	0	7
I am certain I can figure out how to do the most difficult science class work.	2	1	5
My friends and family recognize me as a scientist.	5	0	3
Total	14	14	28

The SIEVEA data from Table 4.2 was used for identifying any changes in students' SI. The aggregate results showed 14 positive changes, 14 negative changes and 28 no changes in students' answers to SIEVEA items after completing the River City project. Even though it may appear that there was no change in totality, going through each item reveals a different picture. All items except two showed either positive net change or no change. In fact, four items had more positive changes than negative ones and one item was neutral (equal number of positive and negative changes). Two items were flagged as exceptions, since the negative changes outnumbered the positive changes for them. These two items were the following: “I am confident I can master the skills taught in my science class” and “I consider science topics very interesting and engaging.”

Interestingly, these two items were more related to students’ motivation than to their science identities. Indeed, “I am confident I can master the skills taught in my science class” reflects students’ expectancy in succeeding in science, whereas “I consider science topics very interesting and engaging” deals with how students value science. Therefore, if these two items are excluded, the remaining results show a positive shift in students’ science identities, especially for the item “My friends and family recognize me as a scientist,” which is a strong indicator for science identity.

Next, it was worthwhile to explore why two flagged items showed a negative shift during the River City project. In other words, this generated a new question not present in the research questions’ list: “What was the reason behind the negative shift in students’ motivation?” In order to answer this question, the SIEVEA data was linked to the SIQ data for conducting a cross-analysis between two instruments. Here are the exact steps used for this fine-grained qualitative analysis:

- The two survey items that caused a negative shift in students’ motivation were flagged as follows:
    - I consider science topics very interesting and engaging (SIEVEA Pre-test, Item 5 and Post-test, Item 9)
    - I am confident I can master the skills taught in my science class (SIEVEA Pre-test, Item 4 and Post-test, Item 4)

- Two SIQ questions related to the SIEVEA's two items were identified:
  - Describe your experiences and feelings about the River City project (SIQ Post-test, Q1)
  - Do you think you will feel more confident in your abilities in science classes after your experiences in the River City? (SIQ Post-test, Q7)
- SIEVEA items were linked to SIQ questions as follows: 1) item "I consider science topics very interesting and engaging" was linked to question "Describe your experiences and feelings about the River City project," 2) item "I am confident I can master the skills taught" was linked to question "Do you think you will feel more confident in your abilities in science classes after your experiences in the River City?" Students' answers to two SIQ questions were analyzed in order to get more insights into the negative shifts of two SIEVEA items.

Tables 4.3 & 4.4 summarize students' responses on SIEVEA items and their corresponding SIQ questions.

Table 4.3
SIQ (Q1) vs SIEVEA (Item 5 Pre-Test and Item 9 Post-Test)

Student		Response
Lea (S1)	Pre-Test (Agree) Post-Test (Agree)	The River City project was a new and interesting experience to me. Although, I already knew a lot of things it wanted you to learn, it was a creative idea in the first place. I had difficulty managing it as it would often erase my data collected.
Shanaia (S2)	Pre-Test (Neither Agree nor Disagree) Post-Test (Disagree)	River City was somewhat of an interesting experience. It was sometimes difficult to keep up with but it soon became easier to make observations and identify part of the problems and some of its sources.
Samad	Pre-Test (Strongly Agree)	The River City project helped me broaden my science skills. River City helped me make predictions/

(S3)	Post-Test (Agree)	hypothesis about how people in the city were getting sick. I eventually found out that mosquitos were causing the sickness in the city. I feel that participating in the River City project helped me make better predictions about what was causing the problem in River City. I felt like a scientist when playing the River City game because I was making predictions, inferences, creating hypothesis, drawing conclusions and collecting data from River City and also interviewing people.
Tom (S4)	Pre-Test (Strongly Agree) Post-Test (Strongly Agree)	The River City project was interesting. Like anything it didn't go flawlessly. There were problems that I faced which stopped me from progressing. The actual project had an interesting concept. But, since we weren't all organized, we ended up working individually rather than as a group.
Tanashia (S5)	Pre-Test (Agree) Post-Test (Disagree)	Everyday felt very repetitive. It was fun at first but then got boring since each day there were so many questions. For July my map and notebook did not pop up.
Jose (S6)	Pre-Test (Neither Agree nor Disagree) Post-Test (Agree)	From my experience and feelings about the River City project I felt that there were a good number of little bugs such as the whole right side where the notebook supposed to be is just white and I also got stuck in door a good number of times. Overall, I feel that the project was pretty entertaining most of the time, but I had a couple boring moments. The game was easy to understand most of the time and I'm glad for doing it.
Juan (S7)	Pre-Test (Agree) Post-Test (Neither Agree nor Disagree)	In my opinion the River City project was a good experience. I was able to act and be a scientist for a couple of days. I was able to make observations and get a lot of info for a problem that was affecting River City in many big ways. This project gave me a little different perspective of science and scientist.
Jessica (S8)	Pre-Test (Agree) Post-Test (Neither Agree nor Disagree)	The game itself was fine; it wasn't too difficult to understand. The only issue I had to deal with was that sometimes when I logged into the game my notebook/ map was a blank page so I couldn't work on the day even though I had the software downloaded. The game when I could get it to work wasn't difficult, following the instructions I was able to complete some if it and everything was simple and easy to understand.

Table 4.4
SIQ (Q7) vs SIEVEA (Item 4 Pre-Test and Item 4 Post-Test)

Student		Response
Lea (S1)	Pre-Test (Agree) Post-Test (Agree)	My experience in River City felt that of my 6th grade science class because I had already learned those topics, simply. My science classes recently are a little more challenging than just observations and inferences.
Shanaia (S2)	Pre-Test (Strongly Agree) Post-Test (Neither Agree nor Disagree)	No, this actually made me less confident about my abilities in science class.
Samad (S3)	Pre-Test (Agree) Post-Test (Agree)	I will feel more confident in my abilities in science classes after my experiences I the River City project because I broadened and expanded my knowledge in science by solving the mystery in the River City project.
Tom (S4)	Pre-Test (Strongly Agree) Post-Test (Agree)	In a way maybe. River City showed me that science processes can take a long time and it might help me be more patient one day.
Tanashia (S5)	Pre-Test (Strongly Agree) Post-Test (Agree)	No.
Jose (S6)	Pre-Test (Agree) Post-Test (Neither Agree nor Disagree)	Yes, I actually do feel more confident because I've learned many new things from River City.
Juan (S7)	Pre-Test (Agree) Post-Test (Agree)	Yes, I feel a little more confident in my abilities because it taught a lot on science.
Jessica (S8)	Pre-Test (Agree) Post-Test (Neither Agree nor Disagree)	No, I struggle with science and the project hasn't helped me with my confidence.

The qualitative analysis of the data in Tables 4.3 & 4.4 helped in understanding why there was a negative shift in students' motivation.

Shanaia (S2) experienced some difficulties in the River City project; "It was sometimes difficult to keep up with," which could have negatively influenced her interest in science topics, even though she found the project to be "somewhat of an interesting experience." Likewise, she thought that the River City experience made her "less confident" about her "abilities in science class." This perfectly explains why her answers on the SIEVEA item "I am confident I can master the skills taught in my science class" changed from Strongly Agree (pre-test) to Neither Agree nor Disagree (post-test). Overall, Shanaia experienced a significant drop in her motivation in science.

Tanashia (S5) did not enjoy the River City project. Her comments like "Everyday felt very repetitive" and "It was fun at first but then got boring since each day there were so many questions" clearly show that she was not thrilled with this project. Also, when asked if she would feel more confident in her abilities in science classes after her experiences with River City, her answer was a laconic, but firm, "No." Tanashia's lack of enjoyment with the project clearly affected her motivation: on pre-test and post-test SIEVEA items, her responses experienced a drop.

Jessica (S8) also experienced some difficulties with the River City project; "sometimes when I logged into the game my notebook/ map was a blank page so I couldn't work on the day." She also responded to SIQ question Q7 with: "No. I struggle with science and the project hasn't helped me with my confidence." Here again, it seemed that her experience with the project negatively influenced her science motivation.

It seemed that Lea (S1) already knew lots of material that was covered in the project. She mentioned in her answers, "I already knew a lot of things it wanted you to learn" and "I had already learned those topics." Overall, she found science topics interesting before and after participating in the River City project. She had a positive experience with the project. Consequently, it was not surprising that there was no change in her science motivation.

Samad's (S3) experience with the River City project was positive. He listed how it helped him to practice his skills in conducting observations, formulating hypothesis, making inferences, collecting data based on virtual labs, and creating conclusions. He also enjoyed interviewing River City's residents. Nevertheless, Samad's survey data showed a slight decrease in his interest in science topics. His "Strongly Agree" answer on the pre-test changed to "Agree" on the post-test. For the second question, Samad mentioned that the River City project "expanded and broadened" his scientific knowledge. However, his survey data showed no change in his beliefs; he chose "Agree" on both pre- and post-tests.

Tom (S4), Jose (S6) and Juan (S7) all had positive experiences with the River City project. They considered the project interesting, despite some technical and organizational flaws. Their survey data indicated no change or a slight change in their science motivation. For the second question, Tom did not express strong confidence regarding his abilities. His justification was that he did not realize that the science process requires patience. Maybe Tom was not a patient individual and wanted to solve the problem quickly. However, after spending time in the project, he realized that science did not work that way. Tom wrote, "River City showed me that science processes can take a long time and it might help me be more patient one day." Tom's "maybe" was reflected in the drop of his confidence in mastering science skills: his "Strongly Agree" answer changed to "Agree" after the project was complete.

To summarize, although there was no noticeable change in students' SI in general, the fine-grained analysis showed a possible but not significant change in SI and even some drop in SI and motivation. This drop was quite pronounced in students' motivation in science rather than in their science identities. Delving deeper in this unexpected negative change and bridging the SIEVEA and SIQ data revealed that the negative change occurred due to three students' responses. It is safe to assume that not all students contributed to the drop in motivation in science, but rather three students: Shanaia, Tanashia and Jessica. These three students' negative experiences with the project were behind of this unexpected phenomenon. The analysis of these students' written statements showed that Shanaia's concern was with the difficulty of the task, Tanashia brought up the amount of the work and Jessica mentioned her experiences with technical difficulties. "Struggling with the task" is an appropriate phrase describing all three students' experiences. Since these students struggled and had difficulties either with the quality or with quantity of the work, they lost their motivation toward the project.

SIQ questionnaire – students' science identities. Just like the SIEVEA, students completed the SIQ twice: before starting the project (pre-test) and after finishing the project (post-test). The SIQ was paper-and-pencil and students took it in their science classroom. Each SIQ had nine questions with the following composition (see Appendix B):

1. The last question, Q9, was identical for both pre- and post-test. It asked, "In what ways do you see yourself as a science person?" To code student answers, ones or zeros were assigned to each answer: one was used to indicate an affirmative answer, whereas zero was used to show a negative answer. Then, the pre-test and post-test codes were used for comparing changes in students' science identities before and after the project.
2. Three questions, the pre-test questions #3, #5 and #6 and the post-test questions #4, # 6 and #7, were about the same concern but were phrased differently. The analyses of these questions were focused on the change of students' SI, provided it occurred.
3. Lastly, the remaining five questions were very specific for the pre-test and post-test. Pre-test questions #1, #2, #4, #7 and #8 were analyzed for gathering additional information regarding students' grades, motivation toward science, the reason for joining to the River City project and their previous experiences related to science. Post-test questions #1, #2, #3, #5 and #8 were treated similarly. These questions were specifically evaluating and inquiring students' perceptions regarding the River City project.

Table 4.5 summarizes how students responded to the pre-test and post-test question #9: "In what ways do you see yourself as a science person?" As was mentioned before, the answers were one- and zero-coded for positive and negative responses, respectively.

Table 4.5
Student Responses to SIQ Question Q9 (Pre- and Post-Test)

Student		Response	Comments
Lea (S1)	Pre-Test (0)	I do not see myself as a science person at all. Later in life I hope to succeed in the fashion industry, which does not approach the scientific ground, really	Lea did not consider her a science person either before or after the project
	Post-Test (0)	I do not see myself as a science person at all.	
Shanaia (S2)	Pre-Test (1)	I see myself going into my dream job of being a psychologist/psychiatrist which fits under the field of science. However, I want to also enter the field of business.	Shanaia had mixed feelings about being a science person: she was considering about entering the science field as well as the business field. Later, she changed her mind about science. Possible explanation – since Shanaia had mixed feelings about science, after realizing difficulties and challenges that scientists face, she decided against it.
	Post-Test (0)	I do not see myself as a science person.	
Samad (S3)	Pre-Test (1)	I see myself as a person who is successful in science. I want to have a job that is in the science field, like a biology teacher or a dentist. I hope to go Rutgers SEBS to get a BS in Biology.	Samad wrote about seeing himself as a science person. In the pre-test he mentioned about the jobs he wanted to pursue. Later, in his post-test he mentioned the specific skills that he had learned from the project.
	Post-Test (1)	I see myself as a moderate scientist because I can solve problems if I have enough given information. I can make predictions and hypothesis of what is causing problems, and I can collect data to help me prove my predictions and hypothesis.	

Tom (S4)	Pre-Test (1)	Science has always been a part of my life. Hopefully in the future I can learn more about astronomy and get a job looking at stuff in space.	Like Samad, Tom talked about his dream job during the pre-test, whereas for the post-test he gave general information about seeing himself as a science person.
	Post-Test (1)	I see myself as a person who can do their work as long as everything works for me. I will think things through and ask others for their responses. I always try to get the answer I feel most confident with and the one I could give the best explanation to.	
Tanashia (S5)	Pre-Test (0)	I'm more of a math person than science person.	It is clear from Brook's statements that she did not consider herself as a science person. However, she mentioned science processes such as making observations and inferences in her post-test.
	Post-Test (1)	When taking observations and inferences.	
Jose (S6)	Pre-Test (0)	I don't really think of myself as a science person. I can't be, science can be extremely fun at times and I do feel like a science person when doing labs and such. But, I just don't think of myself as a science person, more like in engineering or electrical person.	Jose mentioned in his pre-test that, when he did labs, he considered science as fun subject and thought about himself as a science person. However, he stated twice that in general he did not consider himself a science person. For his post-test Jose reaffirmed once again that he did not see himself as a science person.
	Post-Test (0)	I do not see myself as a science person in any way.	
Juan (S7)	Pre-Test (1)	I see myself probably working with chemistry and doing plenty of experiments.	Interestingly, Juan considered himself as a science person doing experiments related to chemistry. However, in the post-test he clearly stated that science was not for him and he did not consider himself as a science person. It is clear that Juan did not like science processes such as
	Post-Test (0)	I didn't really see myself as a science person overall. I don't see me doing observations and taking notes on problems the society is having. Science is not my thing and I don't feel that science will ever fit in with me.	

			making observations and taking notes while he did not object with hands on activities like doing experiments.
Jessica (S8)	Pre-Test (1)	Although I highly doubt this will happen, I've always enjoyed astronomy and marine biology so I would someday like to participate in work related to these subjects. Even if it’s just observing others that work in fields related to this subject I'd like to experience what's it like to work there. The ocean and space really interests me because of how little we actually know.	Jessica’s pre-test notes showed that she enjoyed astronomy and marine science. Especially her last sentence, where she mentioned how little is known about the ocean and space, was impressive. Although she did not specifically mention that she considered herself as a science person, her genuine interest in these fields left little doubt that she was a science person.
	Post-Test (1)	I see myself as a science person in the aspect that (were not able to transcribe the word) enjoy some sciences and look to learn more. I do not like all sciences though so if I see myself as a science person in marine biology.	In her post-test note Jessica clearly stated that she considered herself as a science person. However, she did not like all sciences, but was rather interested in marine science.

The data in Table 4.5 shows that two students, Lea (S1) and Jose (S6), did not consider themselves to be science persons, neither before nor after the project. Similarly, three students, Samad (S3), Tom (S4) and Jessica (S8), considered themselves as science persons both before and after the project. Tanashia (S5) did not consider herself to be a science person, but after the project, she felt as though she were a science person.

Only two students, Jenai (S2) and Juan (S7), did not consider themselves to be science persons in their post-test, despite indication of science person’s traits in their pre-test. Jenai’s responses to SIQ Q9 were consistent with her SIQ and SIEVEA results, which indicated lack of interest toward science after participating in the River City project. Shanaia's letters to the mayor did indicate a moderate increase in her science identity traits. However, it appeared that her loss of interest weighed more heavily on her results than anything else.

Juan's case was different. Even though his letters to the mayor indicated some increase in his science identity and he felt more confident about his abilities in science (see Juan's response in SIQ question #7 in Table 4.4), it seemed that he did not really think that science was for him. However, he did enjoy some aspects of doing science, like experiments.

Pre-test questions #3, #5 and #6 were matched to the post-test questions #4, # 6 and #7 in order to perform comparisons of students' answers to these questions. These questions, with their matches, were summarized in Table 4.6.

Table 4.6
Pairing of SIQ Pre- and Post-Test Questions

Pre-Test Question	Post-Test Question
Q3. Do you feel motivated in science classes? Why or why not?	Q4. Did you feel motivated while observing the problems in the River City? If yes, provide a specific example.
Q5. Do other students and teachers recognize your abilities in science?	Q6. Do you think other students and teachers will recognize you as a science person after you participated in the River City project? Why and why not?
Q6. Are you confident in your abilities in science classes?	Q7. Do you think you will feel more confident in your abilities in science classes after your experiences in the River City?

The same coding schema used for analyzing student responses to Q9 was used for comparing/contrasting the above-mentioned three question pairs. The detailed analysis of responses to pre-test question Q3 and post-test question Q4 were summarized in Table 4.7.

Table 4.7
Student Responses to SIQ Questions Q3 (Pre-Test) and Q4 (Post-Test)

Student		Response	Comments
Lea (S1)	Pre-Test (1)	I feel motivated in my science classes to become more aware of issues that science helps explain. Along with finding further discoveries of what really goes on in my body why things happen, how and help others realize these points also.	Lea's both pre- and post-test results indicated motivation toward science. Moreover, it seemed that records from hospital admissions made Lea intrinsically motivated.
	Post-Test (1)	I felt like it was very easy to understand what the problem in River City was. But at first, records from hospital admissions made me more curious.	
Shanaia (S2)	Pre-Test (0)	I sometimes feel motivated but there are many times where I'm just lost with classwork or tests. Also, I feel as though I'm in a dare when my teacher is going (not finished sentence)	Although Shanaia mentioned that she sometimes felt motivated in her science classes, she added that she was lost in her tests and classwork and wanted to put comments about the teacher but did not finish the sentence. Also, in her post-test, she clearly stated that River City did not help her with the motivation. She stated that she felt lost since it was a new area for her.
	Post-Test (0)	No, I didn't, because I felt a little lost being in River City; it was a new area for me.	
Samad (S3)	Pre-Test (1)	I do feel motivated in science classes because I know that in the future, the material or curriculum taught in class will be beneficial for me in the future. So, that's why I feel motivated in science class.	Samad clearly stated in his pre-test that he felt motivated in his science class. Moreover, he explained why by stating that it "will be beneficial for me in the future." He showed the same enthusiasm in his post-test.
	Post-Test	I felt motivated when solving problems in River City because I	

	(1)	actually wanted to solve the enigma of what was causing the problem. When I was interviewing people from River City, I started to understand what was causing the problem in River City.	He claimed that River City project kept him motivated because "I actually wanted to solve the enigma of what causing the problem." Samad's statement is another example of intrinsic motivation demonstrated by the students.
Tom (S4)	Pre-Test (1)	I feel motivated in every class I attend or any activity I am part of. The reason I feel like this is because of my family. My family is from Europe and when we came here we had nothing. I was the one that had to learn Engish and do the best in school. This fact always makes me want to work harder.	Tom commented that he felt motivated in every class because his immigrant background motivated him to work hard in order to succeed and help his family out. Tom also felt motivated during the River City project. Interestingly, he connected the idea of motivation to his daily life by stating that" I felt motivated because it might help me one day."
	Post-Test (1)	Yes, I did feel motivated solving the sickness problem in River City. I wanted to figure out what was causing their … and how can I fix it. I felt motivated because it might help me one day.	
Tanashia (S5)	Pre-Test (1)	Yes, my lab partners help me and when I'm alone I feel motivated	In her pre-test, Tanashia indicated about feeling motivated in science. However, in her post-test she simply said that she did not feel motivated "because everything took so long to do and I got bored."
	Post-Test (0)	No, because everything took so long to do and I got bored.	
Jose (S6)	Pre-Test (1)	There are times I feel motivated in science classes but that is usually times when I do hands on activities. I rarely get motivated, to be honest, when taking notes in class because it is a bit boring to me. I generally feel motivated when what we do in science class feels fun to me and that is an exciting experience.	Jose explicitly stated that only those activities, which he thought were interesting, made him motivated in science. In his post-test, he confirmed by claiming that there were "only certain number of problems" that motivated him. As an example, he mentioned secret assignments as "cool and interesting."
	Post-Test (1)	To be honest, there were only certain number of problems that interested in River City. One	

		example would be the secret assignments we did, I thought there were interesting and cool.	
Juan (S7)	Pre-Test (0)	No, not really. I don't feel motivated on learning it or doing my work either. Science doesn't really suit me that much, I find science boring.	Juan stated in his pre-test that science was not for him and he did not find it motivating or interesting. He maintained the same belief in his post-test.
	Post-Test (0)	No, I did not feel motivated to solve the problems in River City. There were too many questions, and sometimes I found the questions repetitive.	
Jessica (S8)	Pre-Test (1)	I feel motivated to learn about some different areas of science. I really like marine biology and astronomy and feel motivated to learn about these subjects. Other than those I don't really feel motivated.	Jessica's motivation in science was limited to certain subjects: astronomy and marine biology. Her statements showed that she was both extrinsically (extra credit) and intrinsically (assignments were fun and secretive) motivated.
	Post-Test (1)	I feel motivated because at the end of the day I knew I'm getting extra credit. But, also I felt motivated to do the secret assignments because they kind of fun and secretive.	

The students' responses to SIQ pre-test question Q3 and post-test question Q4 highlighted the fact that most students were motivated in science before and after participating in the River City project. Shanaia and Juan were the obvious exceptions. They were not motivated both before and after their participation in the project. This was consistent with their answers and responses to other SIEVEA items and SIQ questions. Tanashia showed a drop in her motivation after the project. She claimed, "everything took so long to do and I got bored." Again, this response was expected, given her answer to SIQ question Q1: "Everyday felt very repetitive. It was fun at first but then got boring since each day there were so many questions."

The analysis of SIQ's pre-test question Q5 and post-test question Q6 were included in Table 4.8.

Table 4.8
Student Responses to SIQ Questions Q5 (Pre-Test) and Q6 (Post-Test)

Student		Response	Comments
Lea (S1)	Pre-Test (1)	Teachers recognize my abilities in science as just regular, average relationship with it. Other students though think I'm super smart and excel at it - when that is not the case.	Lea thought she was recognized by her peers as an able person in science but not by her teachers. However, she herself did not think she was capable in science. In her post-test, she claimed that she was not recognized by others because she did not "place her best interest in science."
	Post-Test (0)	Because I do not place my best interest in science, no one will recognize me as a science person.	
Shanaia (S2)	Pre-Test (1)	Yes, I've learned that students and teachers recognize my abilities in science.	Shanaia thought that others recognized her as a science person. However, after the project she felt differently because of her struggles with the project.
	Post-Test (0)	No, they wouldn't, because I didn't find it interesting and I was struggling with it.	
Samad (S3)	Pre-Test (1)	I believe that my teachers and fellow classmates do recognize my abilities in science, hopefully they do. Hopefully, my teachers understand that science is my passion and its something I enjoy doing it.	Prior to the project, Samad thought his teachers and classmates recognized him as a science person. After project's completion, his belief in this respect got stronger.
	Post-Test (1)	I hope people like students and teachers recognize that I'm a science person after I participated in the River City project because I participated in the River City project to show students and teachers that I had great knowledge in science and that I was interested in science. I proved that I am a science person by solving the mystery what was causing the problem in River	

		City.	
Tom (S4)	Pre-Test (1)	Many students feel like I am the smartest kid in my grade. They often way this because of my work either grades. My 7th and 8th grade science teachers also thought I was … of the best students and … I had great abilities in science.	Tom felt that his classmates and 7th and 8th grade teachers thought he had strong abilities in science. His answer after participating in the River City project showed some confusion. It seemed that he thought that his participation in the project would not make any difference on how people perceive him as a science person because people would not think it was important.
	Post-Test (1)	No, they can't because no one would seem to care. The only people that would recognize would be people who already see me as a science person.	
Tanashia (S5)	Pre-Test (1)	My science teachers believe in me and see my capability.	Tanashia believed her science abilities were recognized. Although she provided a short answer for her post-test, it still indicated recognition by others.
	Post-Test (1)	Yes, because this is a science game	
Jose (S6)	Pre-Test (1)	I don’t really think that my teachers recognize my abilities in science. On the other hand, I am pretty sure some of my friends recognize my abilities in science because they ask for help at times and are impressed by my grades.	Jose believed he was recognized by his friends regarding his science abilities. He did not think his participation in the River City project would make any difference since nobody would know about it.
	Post-Test (0)	Not really, because nobody really knew I did this and I doubt I'll use what I learned in the future.	
Juan (S7)	Pre-Test (0)	I'm not sure if other people recognize my abilities at all.	Juan did not think that he was recognized by others. He did not feel the River City project would make any difference because it nobody could see what he was doing.
	Post-Test (0)	No, all I did was just go into a game and try to solve a problem. If I actually could have it done it physically then people would most likely look at me different.	
Jessica (S8)	Pre-Test (0)	I don’t think other students and teachers recognize my abilities in science. I just do my work and hope to get a good grade but no one recognizes me for it.	Jessica did not think she was recognized as a science person. She also felt nothing would change after the River City project since

Post-Test (0)	No, many teachers and students don't know about my participation in the game so I don't believe they would see me as a science person. Even if I did I don’t think I'd be considered a science person because they would see this as another school assignment.	her teachers and peers would not know about her participation in the project.

Responses to SIQ’s pre-test question Q5 and post-test question Q6 revealed a flaw in how question Q6 was worded: “Do you think other students and teachers will recognize you as a science person after you participated in the River City project? Why and why not?” Since the River City project was online, some students thought it would preclude their teachers and peers from knowing about their participation in the project. Because of this, these students gave a negative answer to this question. These students were Jose, Juan and Jessica. From the remaining students, Lea was ambivalent about how she was viewed by her teachers and peers. She continued thinking she was not a science person after the project. Shanaia did not think she was a science person either, but her reasoning was tied to her lack of interest in the project.

The summaries of SIQ’s pre-test question Q6 and post-test question Q7 were included in Table 4.9. According to the data, the participation in the River City project either maintained or improved students’ confidence in their abilities in science classes, except for three students: Shanaia, Tanashia and Jessica. Shanaia and Tanashia had a drop in their confidence, whereas Jessica was not confident both before and after the project. These students’ responses to these questions aligned well with their answers to other questions.

Table 4.9
Student Responses to SIQ Questions Q6 (Pre-Test) and Q7 (Post-Test)

Student		Response	Comments
Lea (S1)	Pre-Test (1)	I have a normal confidence in my scientific abilities. I am not too bold with it, but I pull it out of the bag once in a while.	In both pre- and post-tests Lea indicated confidence in her science abilities.
	Post-Test (1)	My experience in River City felt that of my 6th grade science class because I had already learned those topics, simply. My science classes, because I had already learned those topics, simply. My science classes, recently are a little more challenging than just observations and inferences.	
Shanaia (S2)	Pre-Test (1)	I am confident enough to know I can receive a B in a science class.	Shanaia had a drop in her confidence after the River City project.
	Post-Test (0)	No, this actually made me less confident about my abilities in science class.	
Samad (S3)	Pre-Test (1)	I am confident in my abilities in science classes because I enjoy science and get good grades in my science classes. I know that sometime in the future, I will end up doing a job that is in the science field.	Samad was positive about his abilities in science classes, which he mentioned in his both pre- and post-tests. Moreover, he clearly stated in his post-test that the River City made him feel more confident about science.
	Post-Test (1)	I will feel more confident in my abilities in science classes after my experiences I the River City project because I broadened and expanded my knowledge in science by solving the mystery in the River City project.	
Tom (S4)	Pre-Test (1)	Yes, I am confident in my abilities in science. I work hard, remember things easily and I can work on my own.	Tom stated that he felt confident about science in his pre- and post-tests. Furthermore, Tom mentioned that the project might help him to become more patient while working on science processes.
	Post-Test (1)	In a way maybe. River City showed me that science processes can take a long time and it might help me be more patient one day.	
Tanashia (S5)	Pre-Test (1)	I am very confident in my abilities.	Tanashia stated that she was confident in her abilities. However, she felt that the project made no difference for her.
	Post-Test (0)	No.	

Jose (S6)	Pre-Test (1)	I feel fairly confident in most subjects in science class. There are times though where I have been very confused with a certain thing which we learned about and has caused a bad grade.	Jose's pre-test shows that he felt confident in science even though several times he felt confused about certain topics. In his post-test, Jose clearly stated that he felt more confident after the River City project since he learned many new things.
	Post-Test (1)	Yes, I actually do feel more confident because I've learned many new things from River City.	
Juan (S7)	Pre-Test (1)	Yes, I am confident in some of my abilities. I am stronger when it comes to writing down open-ended, rather than choosing multiple-choice.	It seemed that Juan was confident in his abilities before the project. The project seemed to improve his confidence.
	Post-Test (1)	Yes, I feel a little more confident in my abilities because it taught a lot on science.	
Jessica (S8)	Pre-Test (0)	With something I'm confident because I understand it. Other things I think I have to study more and review my notes to understand. I think biology is the class I'm least confident in but I hope to improve	Jessica was not confident about her abilities before the project. It seemed that the project did not help her in building more confidence.
	Post-Test (0)	No, I struggle with science and the project hasn't helped me with my confidence.	

The remaining SIQ pre-test and post-test questions provided the researcher with more data, such as information regarding students' grades, their reasons for joining the project, and so on. Based on this data, it became evident that all student participants had A's and/or B's in their science class. None of the participants had a grade lower than B (pre-test Q4). In addition, all students, except for one, had positive experiences in their previous science classes (pre-test Q1). However, the data revealed something else that was not expected: blaming teachers for not teaching well, not doing labs and not making the subject fun for students. Some students liked their science classes in elementary/middle schools, but not in high school. The students' voices were captured in Appendix D.

Given the findings discussed above, it can be concluded that a) all research participants had either A's or B's in their science class, b) all students, expect for one, had positive experiences in their previous science classes, c) some students liked their elementary/middle school science classes but not their high school science classes, d) some students blamed their teachers for not making science learning a fun and interactive experience, e) most students' confidence in their abilities in science classes were either improved or remained unchanged after the River City project, f) most of the students were motivated before and after their participation in the River City project.

Discussion. During the River City project, students used 21st century skills and technology to come up with sound explanations of the root causes of problems and recommended viable solutions. The project made students realize that one can be a scientist or a science person without wearing a white gown or necessarily working in a lab. In contemporary cognitive psychology, the "learning-by-doing" approach is utilized mostly "with a combination of abstract instruction and concrete illustrations" (Anderson, Reder, & Simon, 1996, p.8). The River City project fits perfectly with this characterization.

It was hypothesized that activities like River City would help them in becoming better decision makers. Additionally, because of this study, a shift in students' identities towards more scientific identities was anticipated. It should be noted that construction of persistent identities can take several years and happens across many contexts (Horn, 2008). Nevertheless, fine-grained qualitative analysis of River City data collected over two weeks indicated some changes in students' science identities.

Indeed, the analysis of students' letters to the mayor of River City showed some improvements in their science identity competence and performance scores. After the data was analyzed, refined and interpreted, it became apparent that students' science identities were present in their letters to the mayor of River City. Moreover, the pre- and post-letters' mean scores and individual scores showed some indications of increased science identities. This result strengthened the case for implementing web-based, constructivist teaching strategies and supported the idea that science identities were not stagnant, but rather that they evolved (Moje et al., 2007).

Psychometrically validated survey instrument SIEVEA was also used for scrutinizing high school students' science identities. The results of the SIEVEA provided more evidence of the positive shift in students' science identities due to the River City project. This shift was particularly pronounced for the item "My friends and family recognize me as a scientist" indicating a significant change in students' science identities. Additionally, students' responses to the SIQ questionnaire items showed a strengthening of science identities for the majority of students. In particular, the analysis of students' answers to SIQ pre-test question "Are you confident in your abilities in science classes? " and post-test question "Do you think you will feel more confident in your abilities in science classes after your experiences in the River City?" demonstrated that the participation in the River City project either preserved or enhanced students' confidence in their abilities in science classes.

Although the qualitative analyses of this study indicated some change in students' science identities, it should be noted that this result was not conclusive due to the small sample size, the short duration of the project and measurement errors. The majority of previous research found science identity changes in the context of longitudinal studies. It is arguable that students' science identities could have noticeably changed over the two weeks period.

According to Grinell (2013), science learning can happen in three different ways: through textbooks, linear models or as a result of daily practice. Grinell argued further that, although the textbook and linear model provide learners with the theory of science and lots of facts, they do not help much with understanding the actual science practice and what takes place in the field or in the lab. By employing science practices and acting as scientists, students had a much better learning experience in the River City project than during textbook learning. This was probably one of the reasons why most of the students mentioned in their essay that the River City was a new and interesting science learning experience. The River City taught students how science actually works. However, the SIEVEA data showed drop in some students' motivation. Further analysis, using students' answers to the SIQ questionnaire, helped in explaining this phenomenon. It turned out that this decrease of motivation was due to three students, Shanaia, Tanashia and Jessica. Their responses to SIQ questions revealed that these students either lacked interest in the project or experienced some difficulties during it.

During the River City project students were introduced to the questioning techniques; they worked on the question-stems and were asked to practice constructing meaningful, good quality questions. Due to the time and resource limitations of the project, there was not enough data collected to do in-depth analysis on how students' questioning would change during the project. Instead, the study explored if question stems would help students in constructing meaningful, good quality questions. The collected data showed that almost all student-generated questions were meaningful: twenty-four questions out of twenty-seven.

Conclusion. In the River City project students tackled the real-world problems by exploring the underlying mechanism of the scientific phenomena, rather than by merely relying on well-known scientific facts. After collecting information and data about health problems plaguing River City, students had an opportunity to put these various scattered pieces of evidence together and propose solutions. According to the qualitative data analyses, as the students progressed through this project, their science identities changed to some extent. However, this result was not conclusive, and it seemed unrealistic that any noticeable changes to students' science identities could have happened during two weeks.

Indeed, students' questionnaires, used in lieu of interviews, indicated that some students' science identities and motivation in science did not change or changed slightly after participating in the River City project. The observable changes were not significant and could be attributed to measurement errors. Interestingly, there was an unexpected drop in science motivation of three students. A fine-grained analysis of this phenomenon uncovered that these three students experienced certain technical or task-related difficulties with the project. It seemed that these difficulties were the primary reason of their motivation drop.

To the best of the author's knowledge, the River City project has always taken place in a classroom environment. It has never been done as an enrichment program or as an out-of-classroom learning experience. Learning experiences outside the classroom are much different from the in-class ones. Indeed, during out-of-classroom learning, the teacher is not around to provide immediate assistance regarding content/solution questions or technical problems. Since students were working on the River City project from home, they took ownership of their learning and had to figure out how to overcome any encountered problems and challenges on their own.

Recommendations for future research. Although the use of Google Docs facilitated students' learning of questioning techniques, it seemed that there was still a room for enhancing this process. For example, despite students having an access to Alison King's question stems, the process of making high quality questions was not as interactive and engaging as the River City activities. The question-stems (King, 1984) were developed for generating high quality questions and thoughtful, elaborated answers during face-to-face collaborative learning. For the current study, these question starters were used for online rather than face-to-face interactions.

In order to make this process more enjoyable and engaging, it is recommended to utilize collaboration scripts (Kollar, Fischer, & Hesse, 2006). The extensive body of literature demonstrated the effectiveness of developed collaboration scripts in various domains where various buttons, schemas, sentence strictures made learning the topic fun, interactive and effective (Baker & Lund, 1997; Guzdial & Turns, 2000). So, it is recommended to create question stems in online collaborative settings, which will not only help students to be trained in questioning techniques, but will also make the process fun and interactive. A specific learning platform for collaboration scripts, called Scripting for Collaborative Online Learning (S-COL), is effective and can be used with variety of content (Wecker, Stegmann, Bernstein, Huber, Kalus, Rathmeyer, Kollar, & Fischer, 2010). Moreover, S-COL offers users flexibility and re-usability in the computer-supported collaborative learning (CSCL) environment. Opportunities for future research include designing a learning platform similar to S-COL, which will allow students to utilize question stems already embedded in the technical frame for constructing higher-order thinking questions.

The River City allowed students to be engaged in a problem-based learning context, where they were intrigued to solve the problem of a mystery disease plaguing River City. As research stated in the past, problem-based learning (PBL) is a self-directed learning, where learners can apply their acquired knowledge to new situations (Hmelo & Lin, 2000). This suggests that knowledge transfer may take place because of activities similar to the ones that happened in the River City project. How often transfer of knowledge happens is still subject of many discussions and disagreements (Schwartz, Bransford & Sears, 2005). However, it is widely accepted that transfer is a gradual process of growing, categorizing and organizing knowledge resources and extending them to other situations (Wagner, 2006). Research claimed that after students learn science in a particular context; their knowledge can be transferred to a different context (Gilbert, Bulte, & Pilot, 2011). Other studies (Bransford, Brown, & Cocking, 2000) supported Gilbert et al. (2011) findings.

It was expected that knowledge-change processes would take place among students doing the River City project, while also transferring their skills in other contexts.

However, transfer of knowledge was not investigated during the River City research project even though it was assumed that students should be able to transfer problem-solving skills they learned during this project to other discipline areas. Therefore, it is recommended to examine the presence of transfer while students are engaged in virtual worlds activities similar to the River City.

Lastly, one of the predicted outcomes of this project was the development of students' inquiry skills through the process of designing and conducting real world investigations in an online environment. The ICAP framework proposed by Chi and Wylie (2014) suggested that cognitive engagement activities, including activities in the interactive computer-based learning environment, could be classified as passive, active, constructive and interactive, depending on the specific tasks that students are engaged in. The River City project falls under both interactive and constructive cognitive learning categories, because students had many opportunities to exchange ideas (Rafal, 1996), ask questions and answer each other's questions (Webb, 1989), including self-questioning (Kramarski & Dudai, 2009). Also, longitudinal studies conducted with high school students found that real-world problems and inquiry-based experiments that offer more than one solution, indeed promote the development of critical thinking (Miri, David, & Uri, 2007). Although the development of inquiry skills was hypothesized, it was not researched for the current study. Therefore, it is recommended to examine students' newly gained inquiry skills resulted from their participation in the River City project.

References

Aghekyan, R. (2015). Questioning Patterns during Discussions in Collaborative Groups in Socioeconomically Diverse High Schools. *NERA Conference Proceedings 2014.* Paper 6. http://digitalcommons.uconn.edu/nera_2014/6

Andersen, L., & Ward, T. J. (2014). Expectancy-value models for the STEM persistence plans of ninth-grade, high-ability Students: A comparison between Black, Hispanic, and White students. *Science Education*, *98*(2), 216-242.

Anderson, J. R., Reder, L. M., & Simon, H. A. (1996). Situated learning and education. *Educational Researcher*, *25*(4), 5-11.

Anderson, L. W., Krathwohl, D. R., & Bloom, B. S. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives.* Boston, MA: Allyn & Bacon.

Baram Tsabari, A., Sethi, R. J., Bry, L., & Yarden, A. (2006). Using questions sent to an Ask-A-Scientist site to identify children's interests in science. *Science Education*, *90*(6), 1050-1072.

Biswas, G., Schwartz, D. L., Leelawong, K., Vye, N., & TAG-V. (2005). Learning by teaching: A new paradigm for educational software. *Applied Artificial Intelligence*, *19*(3).

Bloom, B. S., Engelhart, M. B., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals* (Handbook 1: Cognitive domain). New York: Longmans Green.

Bradley, M. E., Thom, L. R., Hayes, J., & Hay, C. (2008). Ask and you will receive: How question type influences quantity and quality of online discussions. *British Journal of Educational Technology*, *39*(5), 888-900.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press.

Brill, G., & Yarden, A. (2003). Learning biology through research papers: A stimulus for question-asking by high-school students. *Cell Biology Education*, *2*(4), 266-274.

Bryan, L. A., & Atwater, M. M. (2002). Teacher beliefs and cultural models: A challenge for science teacher preparation programs. *Science Education*, *86*, 821–839.

Burke, P. J. (1980). The self: Measurement requirements from an interactionist perspective. *Social Psychology Quarterly*, 18-29.

Burn, A., Buckingham, D., Parry, P., & Powell, M. (2010). Minding the gaps: Teachers' cultures, students' cultures. In D. E. Alvermann (Ed.), *Adolescents' online* *literacies: Connecting classrooms, digital media, and popular culture* (pp. 183-201). New York: Peter Lang.

Byrne, K. A., Patrick, C. J., & Worthy, D. A. (2016). Striatal dopamine, externalizing proneness, and substance abuse effects on wanting and learning during reward-based decision making. *Clinical Psychological Science*, *4*(5), 760-774.

Carlone, H. B., & Johnson, A. (2007). Understanding the science experiences of successful women of color: Science identity as an analytic lens. *Journal of Research in Science Teaching*, *44*(8), 1187-1218.

Carr, D. (1998). The art of asking questions in the teaching of science. *School Science Review*, *79*, 47–50.

Chang, K. E., Chen, Y. L., Lin, H. Y., & Sung, Y. T. (2008). Effects of learning support in simulation-based physics learning. *Computers & Education*, *51*, 1486–1498.

Chiel, H. J., McManus, J. M., & Shaw, K. M. (2010). From biology to mathematical models and back: Teaching modeling to biology students, and biology to math and engineering students. *CBE-Life Sciences Education*, *9*(3), 248-265.

Chi, M. T., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, *49*(4), 219-243.

Chin, C. (2006). Using self-questioning to promote pupils' process skills thinking. *School Science Review*, *87*(321), 113–122.

Chin, C., Brown, D. E., & Bruce, B. C. (2002). Student-generated questions: A meaningful aspect of learning in science. *International Journal of Science Education*, *24*(5), 521–549.

Chin, C., & Chia, L. G. (2004). Problem-based learning: Using students' questions to drive knowledge construction. *Science Education*, *88*(5), 707-727.

Chin, C., & Kayalvizhi, G. (2005). What do pupils think of open science investigations? A study of Singaporean primary 6 pupils. *Educational Research*, *47*(1), 107–126.

Chin, C., & Osborne, J. (2008). Students' questions: a potential resource for teaching and learning science. *Studies in Science Education*, *44*(1), 1-39.

Chin, C., & Osborne, J. (2010). Supporting argumentation through students' questions: Case studies in science classrooms. *Journal of the Learning Sciences*, *19*(2), 230-284.

Cho, Y. H., Lee, J., & Jonassen, D. H. (2011). The role of tasks and epistemological beliefs in online peer questioning. *Computers & Education*, *56*(1), 112-126.

Choi, I., Land, S. M., & Turgeon, A. (2005). Scaffolding peer-questioning strategies to facilitate metacognition during online small group discussion. *Instructional Science, 33*, 483–511.

Clarke, J., & Dede, C. (2009). Design for scalability: A case study of the River City curriculum. *Journal of Science Education and Technology, 18*(4), 353-365.

Comer, J.P., Haynes, N.M., Joyner, E.T., & Ben-Avie, M. (1996). *Rallying the whole village: The Comer process for reforming education*. New York: Teachers College Press.

Colucci-Gray, L., Camino, E., Barbiero, G., & Gray, D. (2006). From scientific literacy to sustainability literacy: An ecological framework for education. *Science Education*, *90*(2), 227–252.

Costa, J., Caldeira, H., Gallastegui, J. R., & Otero, J. (2000). An analysis of question asking on scientific texts explaining natural phenomena. *Journal of Research in Science Teaching*, *37*(6), 602-614.

Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches.* Thousand Oaks, CA: Sage.

Cuccio-Schirripa, S., & Steiner, H. E. (2000). Enhancement and analysis of science question level for middle school students. *Journal of Research in Science Teaching*, *37*(2), 210-224.

Cuevas, H. M., & Fiore, S. M. (2014). Enhancing learning outcomes in computer-based training via self-generated elaboration. *Instructional Science*, *42*(6), 839-859.

Curtis, D. D., & Lawson, M. J. (2001). Exploring collaborative online learning. *Journal of Asynchronous Learning Networks*, *5*(1), 21-34.

Dede, C., Ketelhut, D., & Ruess, K. (2003). *Designing for motivation and usability in a museum-based multi-user virtual environment*. Paper presented at the American Educational Research Association Conference, Chicago, IL

de Koning-Veenstra, B., Steenbeek, H. W., van Dijk, M. W., & van Geert, P. L. (2014). Learning through movement: A comparison of learning fraction skills on a digital playful learning environment with a sedentary computer-task. *Learning and Individual Differences*, *36*, 101-109.

Dillon, J. T. (1988). The remedial status of student questioning. *Journal of Curriculum Studies*, *20*(3), 197-210.

Dresner, M., de Rivera, C., Fuccillo, K. K., & Chang, H. (2014). Improving higher-order thinking and knowledge retention in environmental science teaching. *BioScience*, *64*(1), 40-48.

Duschl, R., & Grandy, R. E. (2013). Two views about explicitly teaching nature of science. *Science & Education, 22*(9), 2109-2139.

Erwin, T. D., & Sebrell, K. W. (2003). Assessment of critical thinking: ETS's tasks in critical thinking. *The Journal of General Education*, *52*(1), 50-70.

Facione, P. A. (1990). *The California critical thinking skills test-college level technical report 1: Experimental validation and content validity.* Millbrae, CA: California Academic Press.

Foote, C. J. (1998). Student-generated higher order questioning as a study strategy. *The Journal of Educational Research*, *92*(2), 107-113.

Ford, M. (2008). Disciplinary authority and accountability in scientific practice and learning. *Science Education*, *92*(3), 404-423.

Gee, J. P. (2000). Identity as an analytic lens for research in education. *Review of Research in Education*, *25,* 99-125.

Gilbert, J. K., Bulte, A. M., & Pilot, A. (2011). Concept development and transfer in context-based science education. *International Journal of Science Education, 33*(6), 817-837.

Glaser, B. S., & Strauss, A. A. (1967). *The discovery of grounded theory.* New York, NY: Aldine De Gruyter.

Graesser, A. C., Lu, S., Olde, B. A., Cooper-Pye, E., & Whitten, S. (2005). Question asking and eye tracking during cognitive disequilibrium: Comprehending illustrated texts on devices when the devices break down. *Memory & Cognition, 33*(7), 1235-1247.

Graesser, A. C., & Person, N. K. (1994). Question asking during tutoring. *American Educational Research Journal*, *31*(1), 104-137.

Gresalfi, M. S. (2004). Taking up opportunities to learn: Examining the construction of participatory mathematical identities in middle school classrooms. *Unpublished Dissertation*, Stanford, Palo Alto.

Grinnell, F. (2013). Research integrity and everyday practice of science. *Science and Engineering Ethics*, *19*(3), 685-701.

Gross, M. (2001). Imposed information seeking in public libraries and school library media centers: a common behavior. *Information Research, 6*(2), 6-2.

Guzdial, M., & Turns, J. (2000). Effective discussion through a computer-mediated anchored forum. *The Journal of the Learning Sciences*, *9*(4), 437-469.

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

Harris, J. (2005). Our agenda for technology integration: It's time to choose. *Contemporary Issues in Technology and Teacher Education*, *5*(2).

Hatt-Echeverria, B. (2005). *What does it mean to be smart? The figured world of smartness and student achievement.* Paper presented at the annual meeting of the American Educational Research Association, Montreal, CA.

Heilbronner, N. N. (2011). Stepping onto the STEM pathway: Factors affecting talented students' declaration of STEM majors in college. *Journal for the Education of the Gifted*, *34*, 876–899.

Hines, S. M. (2003). *Multicultural science education: Theory, practice, and promise*. New York: Peter Lang.

Hmelo-Silver, C. E., & Barrows, H. S. (2006). Goals and strategies of a problem-based learning facilitator. *The Interdisciplinary Journal of Problem-Based Learning*, *1*(1), 4.

Hmelo-Silver, C. E., & Barrows, H. S. (2008). Facilitating collaborative knowledge building. *Cognition and Instruction*, *26*(1), 48-94.

Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist*, *42*(2), 99-107.

Hmelo, C. E., & Lin, X. (2000). Becoming self-directed learners: Strategy development in problem-based learning. In D. Evensen & C. E. Hmelo (Eds.), *Problem-based learning: A research perspective on learning interactions* (pp. 227-250). Mahwah, NJ: Erlbaum.

Horn, I. S. (2008). Turnaround students in high school mathematics: Constructing identities of competence through mathematical worlds. *Mathematical Thinking and Learning*, *10*(3), 201-239.

Ismail, H. N., & Alexander, J. M. (2005). Learning within scripted and nonscripted peer-tutoring sessions: The Malaysian context. *The Journal of Educational Research*, *99*(2), 67-77.

Jagger, S. (2013). Affective learning and the classroom debate. *Innovations in Education and Teaching International*, *50*(1), 38-50.

James, W. (1890). *Principles of Psychology*. New York: Holt.

Johnson, L. F., Levine, A., Smith, R. S., & Haywood, K. (2010). Key Emerging Technologies for Elementary and Secondary Education. *Tech Directions*, *70*(3), 33.

Källkvist, M., Gomez, S., Andersson, H., & Lush, D. (2009). Personalised virtual learning spaces to support undergraduates in producing research reports: Two case studies. *The Internet and Higher Education*, *12*(1), 35-44.

Kane, J. M. (2012). Young African American children constructing academic and disciplinary identities in an urban science classroom. *Science Education*, *96*(3), 457-487.

Ketelhut, D. J., & Nelson, B. C. (2010). Designing for real-world scientific inquiry in virtual environments. *Educational Research*, *52*, 151–167.

King, A. (1989). Effects of self-questioning training on college students' comprehension of lectures. *Contemporary Educational Psychology*, *14*(4), 366-381.

King, A. (1990). Enhancing peer interaction and learning in the classroom through reciprocal questioning. *American Educational Research Journal*, *27*(4), 664-687.

King, A. (1991). Effects of training in strategic questioning on children's problem-solving performance. *Journal of Educational Psychology*, *83*(3), 307.

King, A. (1992). Comparison of self-questioning, summarizing, and notetaking-review as strategies for learning from lectures. *American Educational Research Journal*, *29*(2), 303-323.

King, A. (1992a). Facilitating elaborative learning through guided student-generated questioning. *Educational Psychologist*, *27*(1), 111-126.

King, A. (1994). Autonomy and question asking: The role of personal control in guided student-generated questioning. *Learning and Individual Differences*, *6*(2), 163-185.

King, A. (1995). Designing the instructional process to enhance critical thinking across the curriculum. Inquiring minds really do want to know: Using questioning to teach critical thinking. *Teaching of Psychology*, *22*(1), 13-17.

King, A., & Rosenshine, B. (1993). Effects of guided cooperative questioning on children's knowledge construction. *Journal of Experimental Education*, *61*, 127–148.

Koepp, M. J., Gunn, R. N., Lawrence, A. D., Cunningham, V. J., Dagher, A., Jones, T., & Grasby, P. M. (1998). Evidence for striatal dopamine release during a video game. *Nature*, *393*(6682), 266-268.

Kollar, I., Fischer, F., & Hesse, F. W. (2006). Collaboration scripts–a conceptual analysis. *Educational Psychology Review*, *18*(2), 159-185.

Kramarski, B., & Dudai, V. (2009). Group-metacognitive support for online inquiry in mathematics with differential self-questioning. *Journal of Educational Computing Research*, *40*, 377–404.

Kuhn, L., & Reiser, B. (2006). *Structuring activities to foster argumentative discourse*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Kwon, K., Hong, R. Y., & Laffey, J. M. (2013). The educational impact of metacognitive group coordination in computer-supported collaborative learning. *Computers in Human Behavior*, *29*(4), 1271-1281.

Lacasa, P., Martínez, R., & Del Castillo, H. (2011). *Adolescent thinking and online writing after the use of commercial games in the classroom.* Retrieved from http://www.digra.org/wp-content/uploads/digital-library/11312.53222.pdf

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. New York: Cambridge University Press.

Levy, F., & Murnane, R. J. (2012). *The new division of labor: How computers are creating the next job market*. Princeton, NJ: Princeton University Press.

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

Lipman, P. (2004). *High stakes education: Inequality, globalization, and urban school reform*. New York, NY: Routledge.

Liu, L., & Hmelo-Silver, C. E. (2009). Promoting complex systems learning through the use of conceptual representations in hypermedia. *Journal of Research in Science Teaching*, *46*(9), 1023-1040.

MacKnight, C. B. (2000). Teaching critical thinking through online discussions. *Educause Quarterly*, *23*(4), 38-41.

Mayo, M. J. (2007). Games for science and engineering education. *Communications of the ACM*, *50*(7), 30-35.

Miri, B., David, B. C., & Uri, Z. (2007). Purposely teaching for the promotion of higher-order thinking skills: A case of critical thinking. *Research in Science Education*, *37*(4), 353-369.

Moje, E. B., Tucker-Raymond, E., Varelas, M., & Pappas, C. C. (2007). FORUM: Giving oneself over to science: Exploring the roles of subjectivities and identities in learning science. *Cultural Studies of Science Education*, *1*(3), 593-601.

Muuss, R. (1996). *Theories of adolescence (6th ed.)*. New York, NY: McGraw Hill.

National Research Council. (1996). *National Science Education Standards.* Washington, DC: National Academy Press.

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

National Science Foundation. (2001). *Characteristics of doctoral scientists and engineers in the United States. Detailed statistical tables.* Retrieved from https://evergreen.lib.in.us/eg/opac/record/17367200

Needham, D. R., & Begg, I. M. (1991). Problem-oriented training promotes spontaneous analogical transfer: Memory-oriented training promotes memory for training. *Memory & Cognition*, *19*(6), 543-557.

Osborne, J., Erduran, S., & Simon, S. (2004). Enhancing the quality of argumentation in school science. *Journal of Research in Science Teaching*, *41*(10), 994–1020.

Osborne, J. (2014). Teaching scientific practices: Meeting the challenge of change. *Journal of Science Teacher Education*, *25*(2), 177–196.

Otfinowski, R., & Silva-Opps, M. (2015). Writing toward a scientific identity: Shifting from prescriptive to reflective writing in undergraduate biology. *Journal of College Science Teaching*, *45*(2), 19.

Oyserman, D., Bybee, D., & Terry, K. (2006). Possible selves and academic outcomes: How and when possible selves impel action. *Journal of Personality and Social Psychology*, *91*(1), 188.

Palincsar, A. S. (1998). Social constructivist perspectives on teaching and learning. *Annual Review of Psychology*, *45*, 345–375.

Pea, R.D. (1993). Learning scientific concepts through material and social activities: Conversational analysis meets conceptual change. *Educational Psychology, 28*, 265–277.

Pearson, P. D., Moje, E., & Greenleaf, C. (2010). Literacy and science: Each in the service of the other. *Science*, *328*(5977), 459–463.

Peters, V. L., & Hewitt, J. (2010). An investigation of student practices in asynchronous computer conferencing courses. *Computers & Education*, *54*(4), 951-961.

Pink, D. H. (2006). *A whole new mind: Why right-brainers will rule the future*. New York, NY: Penguin.

Rafal, C. T. (1996). From co-construction to takeovers: Science talk in a group of four girls. *Journal of Learning Sciences, 5*, 279–293.

Ratcliffe, M. (1997). Pupil decision-making about socio-scientific issues within the science curriculum. *International Journal of Science Education*, *19*(2), 167–182.

Roscoe, R. D., & Chi, M. T. (2008). Tutor learning: The role of explaining and responding to questions. *Instructional Science*, *36*(4), 321-350.

Rosenshine, B., Meister, C., & Chapman, S. (1996). Teaching students to generate questions: A review of the intervention studies. *Review of Educational Research, 66*(2), 181-221.

Rubin, B. C. (2007). Learner identity amid figured worlds: Constructing (in) competence at an urban high school. *The Urban Review*, *39*(2), 217-249.

Ryan, F. L. (1971). *Exemplars for the new social studies: Instructing in the elementary school*. Englewood Cliffs, NJ: Prentice Hall.

Ryoo, K. (2015). Teaching science through the language of students in technology-enhanced instruction. *Journal of Science Education and Technology*, *24*(1), 29-42.

Saavedra, A. R., & Opfer, V. D. (2012). Learning 21st-century skills requires 21st-century teaching. *Phi Delta Kappan*, *94*(2), 8-13.

Scheuer, O., McLaren, B. M., Weinberger, A., & Niebuhr, S. (2014). Promoting critical, elaborative discussions through a collaboration script and argument diagrams. *Instructional Science*, *42*(2), 127-157.

Schlecher, A. (Ed.). (2012). *Preparing teachers and developing school leaders for the 21st century: Lessons from around the world.* Paris, France: OECD Publishing.

Sfard, A., & Prusak, A. (2005). Telling identities: In search of an analytic tool for investigating learning as a culturally shaped activity. *Educational Researcher, 34*(4), 14-22.

Schwartz, D. L., Blair, K. P., Biswas, G., Leelawong, K., & Davis, J. (2007). Animations of thought: Interactivity in the teachable agents paradigm. In R. Lowe & W. Schnotz (Eds.), *Learning with animation: Research and implications for design* (pp. 114-140). UK: Cambridge University Press.

Schwartz, D. L., Bransford, J. D., & Sears, D. (2005). Efficiency and innovation in transfer. In J. P. Mestre (Ed.), *Transfer of learning from a modern multidisciplinary perspective* (pp. 1-51). Greenwich, CT: Information Age.

Schwartz, M. S., & Fischer, K. W. (2006). Useful metaphors for tackling problems in teaching and learning. *About Campus*, *11*(1), 2-9.

Simon, S., & Amos, R. (2011). Decision making and use of evidence in a socio-scientific problem on air quality. In T. D. Sadler (Ed.), *Socio-scientific issues in the* *classroom* (Vol.39, pp. 167-192). Dordrecht, The Netherlands: Springer Netherlands.

Slavin, R. E. (2008). Comprehensive school reform. In T. Good et al. (Eds.), *21st century education: A reference handbook.* Thousand Oaks, CA: Sage.

Steele, C. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, *52*, 613–629.

Swan, A. E., & O'Donnell, A. M. (2009). The contribution of a virtual biology laboratory to college students' learning. *Innovations in Education and Teaching International*, *46*(4), 405-419.

Swanson, C. B. (2004). *Graduation rates: Real kids, real numbers.* Washington, DC: The Urban Institute.

Tabak, I., & Reiser, B. J. (2008). Software-realized inquiry support for cultivating a disciplinary stance. *Pragmatics & Cognition*, *16*(2), 307-355.

Toledo, C. A. (2015). Dog bite reflections - Socratic questioning revisited. *International Journal of Teaching and Learning in Higher Education*, *27*(2), 275-279.

Van Gelder, T. (2005). Teaching critical thinking: Some lessons from cognitive science. *College Teaching*, 41-46.

Varelas, M. (2012). *Identity construction and science education research: Learning, teaching, and being in multiple contexts* (Vol. 35). Springer Science & Business Media.

Venville, G., Adey, P., Larkin, S., Robertson, A., & Fulham, H. (2003). Fostering thinking through science in the early years of schooling. *International Journal of Science Education*, *25*(11), 1313-1331.

Vosniadou, S. (2013). Conceptual change in learning and instruction: The framework theory approach. *International Handbook of Research on Conceptual Change, 2*, 11-30.

Wagner, J. F. (2006). Transfer in pieces. *Cognition and Instruction*, *24*(1), 1-71.

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

Watson, G. & Glaser, E. M. (1980). *Watson-Glaser critical thinking appraisal manual*: Forms A&B. San Antonio, TX: The Psychological Corporation.

Watts, M., Gould, G., & Alsop, S. (1997). Questions of understanding: Categorizing pupils' questions in science. *School Science Review*, *79*(286), 57–63.

Webb, N. M. (1989). Peer interaction, problem solving, and cognition: Multidisciplinary perspectives. *International Journal of Educational Research*, *13*, 1–119.

Wimer, J. W., Ridenour, C. S., Thomas, K., & Place, A. W. (2001). Higher order teacher questioning of boys and girls in elementary mathematics classrooms. *The Journal of Educational Research*, *95*(2), 84-92.

Wecker, C., Stegmann, K., Bernstein, F., Huber, M. J., Kalus, G., Kollar, I., & Fischer, F. (2010). S-COL: A Copernican turn for the development of flexibly reusable collaboration scripts. *International Journal of Computer-Supported Collaborative Learning*, *5*(3), 321-343.

Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity*. Cambridge, UK: Cambridge University Press.

Williams, R. L., Oliver, R., & Stockdale, S. L. (2004). Psychological versus generic critical thinking as predictors and outcome measures in a large undergraduate human development course. *The Journal of General Education*, *53*(1), 37-58.

Yarden, A., Brill, G., & Falk, H. (2001). Primary literature as a basis for a high-school biology curriculum. *Journal of Biological Education*, *35*, 190–195.

Zohar, A., & Dori, Y. J. (2003). Higher order thinking skills and low-achieving students: Are they mutually exclusive? *The Journal of the Learning Sciences*, *12*(2), 145-181.

Zoller, U. (1994). The examination where the student asks the questions. *School Science and Mathematics*, *94*, 347–349.

Appendix A

Science Identities: Competence and Performance, SICAP

Category	**Sub-Category**
Competence	Knowledge of the science content
	Understanding scientific phenomena
	Motivated to understand the world scientifically
	Understanding the scientific method
Performance	Usage of scientific tools
	Fluency with the scientific talk
	Acting as scientist
	Confident interactions in formal and informal scientific settings

Appendix B

SIQ: Pre-Test

You are asked to answer the following questions. Please make your best effort in providing accurate information.

1. Describe your experiences and feelings about any previous science classes you have taken at any grade level. Provide some specific experiences and memories.
2. Do you think you have good understanding of science? What makes you think in that way?
3. Do you feel motivated in science classes? Why or why not?
4. What kind of grades do you make in your science classes?
5. Do other students and teachers recognize your abilities in science?
6. Are you confident in your abilities in science classes?
7. Why have you decided to participate in the River City project?
8. What do you expect to achieve by participating in this project?
9. In what ways do you see yourself as a science person?

SIQ: Post-Test

You are asked to answer the following questions. Please make your best effort in providing accurate information.

1. Describe your experiences and feelings about the River City project.
2. Provide some memorable parts/events of the project that made you feel like a scientist.
3. Do you think the River City helped you in gaining better understanding of how science works? What makes you think in that way?

4. Did you feel motivated while solving the problems in the River City? If yes, provide a specific example.
5. Have your views about science and “doing” science changed after your participation in the River City project?
6. Do you think other students and teachers will recognize you as a science person after you participated in the River City project? Why and why not?
7. Do you think you will feel more confident in your abilities in science classes after your experiences in the River City?
8. Do you believe that it is important to be scientifically knowledgeable? Does it really matter? Justify your answer.
9. In what ways do you see yourself as a science person?

Appendix C

Students' Letters to the Mayor

<u>Lea (S1)</u>

Pre-letter

"Dear Mayor, through this first journey, I've found out that contamination of water maybe a working case in getting your residents of River City sick. And the area of interest is, mainly within the tenements and young children."

Post-letter

"Dear Mayor,

Throughout my time in River City I've learned various of observations, inferences, symptoms, and the geography of your town. I touched every part of the land from the wealthy homes to the dump. Using my environmental health meter, I observed which areas were the cleanest or more contaminated. It varied in areas like the tenements - which read 50% - and the area between wealthy homes and the scenic overlook (about 75% if I recall). But what really called my attention was witnessing toilet water spill into the town's river. Which also, ultimately happened to be the town's only source of drinking water as well.

After conversing with residents, I began to realize that, mainly in tenements, people would get sick afterwards. Also, with the aid of tools, I've even found E. coli in the water, along with carried viruses within mosquitoes in the environment. Therefore, I've made the inference that these are the causes in the townsfolk becoming sick. In order to rid this issue, more medicines need to be available, mosquitoes must be eliminated, and the bog water cleaned."

<u>Shanaia (S2)</u>

Pre-letter

"Seems as though when I ask people what's new, from one person I learn that the mosquitoes are out about. Then the next clue I discover has to do with someone being sick."

Post-letter

"With River City, I have learned about making observations and forming conclusions from what I have seen. The city has been having health problems lately especially within tenements. Also, a lot of the health issues came from the water. The main symptoms were stomachache and sniffles, some people mainly children, have had fevers and chills. Based off of my observations, I have concluded that their must be tainted water that citizens in tenements are drinking which are causing these symptoms. In order to help this problem, those who drink the bug water can boil it first then maybe refrigerate it to have cold water. The other alternative can be avoid bog water and stick to bottled water. Hopefully, the disease spreading around River City can be avoided by all and eliminated from the town. Good Luck!"

<u>Samad (S3)</u>

Pre-letter

Dear Mayor of River City

I found out that not many residents in River City are sick in 1879 season. It seems as though many of the residents get sick during the summer in River City. The problem might just be that many residents are getting mosquito bites and stomachaches in the summer. Sharon Otissaid when I interviewed her that "It's a lot more fun now since there aren't so many stomachaches.

Post-letter

Dear River City Mayor,

I recommend that your separate tenants from the tenements because they are close to the town dump, so they can become sick, ill because they are near the town dump which has the highest number of mosquitoes. When the tenants/ residents that live near the town dump get bites from mosquitoes they spread then with other residents of River City. Overcrowding also plays a major role in spreading the illness because residents are so close to each other.

<u>Tom (S4)</u>

Pre-letter

"So far we know that most of the poor children living in the tenements are getting sick. Symptoms range from fevers to stomachaches. So far kids in the upper class levels aren't experiencing symptoms. Still missing a good chunk of useful information."

Post-letter

In River City, there were dozens of cases of people getting sick. The first few reports showed that the people living in the tenements were the ones who showed symptoms first. After speaking with the children who lived in the tenements, it was clear that something was wrong in that area. One girl stated that her mother told her to stop playing in the local bog. Using the modern technology, I noticed that there was E. Coli in the water. With all this information I can infer that the water is the problem In River City. There might be a chance that the waterways are all connected in River City. If this was the case, then the bacteria from the tenement area could travel by water to different parts of the town. Now, the question is, how do we fix this problem? The first step is to close the bog in River City. Don't let anyone go in the bog or near to it. The second step is to make save the wells and other water sources are not connected to the bug. The last step would be flush everything out. No more harmful bacteria in the water in River City. The hospitalized residents will just have to keep out of contact with healthy individuals and then River City will be healthy once again.

<u>Tanashia (S5)</u>

Pre-letter

Dear mayor,

I discovered that the only way to help people from getting sick is to clean the trash. The people were drinking dirty bacteria water from the dirty streets filled with trash. Clean up the town so the water will be cleaner so people won't be won't be getting sick.

Post-letter

Dear Mayor,

In river city I've learned that many people were coming down with a fever, cough, and diarrhea. The residents were confused with how and why. In river city I discovered two reasons why the people got sick. It could have been carried by mosquitos. The mosquito has the disease and then it bites the people and transmits the disease in the people. Also, the drinking water had bacteria in it so, it made the people sick. The dirty streets and rain water that washed everything up went into the drinking water. To save the people from getting sick, use bug spray and clean up streets so your drinking water isn't dirty.

<u>Jose (S6)</u>

Pre-letter

Dear Mayor,

Hello and I am sorry for being so late with day 2, I had been having problems with River City but I am back on track now. I found out useful information by my residents and just by observing River City. I found out that symptoms from October are not more but fevers, chills, and coughs are now happening. I saw water pollution happening near the tenements with my own eyes. Seems to be that usually kids are getting these past symptoms and when they stopped going in the river the amount of sick people got less but now there are some new symptoms of something. The cold weather could be an inference to why or just pollution in general.

Post-letter

Dear Mayor,

I have experienced and learned much ever since the beginning of River City. I learned how important and useful investigating and observing could help with finding things out. Many of the NPC's around River City gave me a lot of information that helped me with finding out what is causing disease. The different symptoms for each season also made it a bit challenging to figure out at where this disease came from. First, I noticed the cleanliness of each place in River City and I found out that some streets and places were filthy, rivers being polluted with some odd substances and garbage also the bacteria and mosquitoes found in some rivers. From all that I found out using observation and inference the most logical reason for this would be the filthiness of the water/ rivers around River City and the bacteria found in them. Also, the symptoms in the hospital such as diarrhea and such are closely related to symptoms you would get from drinking polluted water. I have truly learned a lot from doing this assignment and it was pretty fun in my opinion.

<u>Juan (S7)</u>

Pre-letter

After my January visit to River City I have gotten a little closer as to why the people are getting sick. I am now coming to realize that the sickness is most likely coming from the tenements and these people are spreading it throughout the city. I need a little more research to find out what is completely going on.

Post-letter

In River City I learned a lot about the city and how everything is being run. When I first arrived I saw how most of the streets in River City were very dirty. But mainly the tenements area was the most dirty. I saw animal waste and overall garbage. The water near the tenements were also dirty. The only clean area from the River City was the wealthy area of River City. The wealthy area was clean on the streets, and also the water was also clean. From my observations I also saw that inside the water there were a lot of sicknesses. I was able to see E-Coli in the water. In the dump area there were a lot of insect, which I think were spreading the sickness in River City. I believe that the way to solve the sickness in River City is by cleaning the streets and getting rid of all the garbage and waste. Also on top of that clean out the rivers in River City. The water is colored and looks green that needs to be cleaned up. By cleaning these areas most insects will leave and the sickness might not be spread like it is being spread. Hopefully with this advice River City citizens won't get sick.

<u>Jessica (S8)</u>

Pre-letter

Dear Mayor Bowman,

During my visit to river city I have discovered that rumor has it this new illness going around has been speculated to be brought by new comers. Several has reported having is a severe cough. This winters sickness isn't as brutal as the previous summers, I can conclude this from research done at the hospital. Although anyone can become ill it seems what exactly is causing this cough.

Post-letter

Dear Mayor,

What I've learned is that in River city disease and illness spreaded mostly in areas of low maintenance such as tenements and the dump. Although these areas are most common for disease, higher maintenance places still have chances of getting disease. Water being distributed to homes should be filtered to prevent any bugs from being transmitting. A way of fixing these issues is by coming up with new and improved medication. Also, they could exam water that is sent to homes for insects and disease because the water was commonly infected with parasites. This is maybe why so many people got sick; clearing the water is a good way to prevent even more sickness. For those who were infected already with the illness. Hospitals should also work hard to keep their utensils cleaned to avoid spreading more disease and make sure to keep patients separated. They should also make sure to avoid catching any disease to avoid spreading it outside spreading it outside of the hospital.

Appendix D

Students' Voices

<u>Lea (S1)</u>

Science every year was always a 50/50 percent subject for me. The good was I always found little interests in each material taught in order to better understand but I never really actually enjoyed it. In middle school, labs were always my favorite for better judgment, but those are scarce now. Therefore, more to (was not able to read) in to learn.

<u>Shanaia (S2)</u>

Science wasn't the best for me especially in the 7th grade. My educator would try to "teach" my class Biology but I don't think she understood that we were young students learning. My teacher would express her lessons vaguely but in science terms that many of us wouldn't comprehend. Then to "prepare" us for our final exam, she gave us a packet on Genetics expecting us to know what to do when we never covered that section throughout year.

<u>Samad (S3)</u>

I have always loved science classes because we always did fun science experiments. In elementary school science class was basically recess time for me because I had a lot of fun planting and doing other fun science experiments.

<u>Tom (S4)</u>

Science was always one of my favorite classes in school. My teachers always liked me being in their class. Personally I feel most comfortable with science teachers. They are the teachers I often talk to after school or out of class.

<u>Tanashia (S5)</u>

I enjoy doing labs. I like to watch things happen.

<u>Jose (S6)</u>

I believe that I have an average understanding of science. I know that I have a lot more to learn in science but I will that I know enough for my age because of my grades in class.

<u>Juan (S7)</u>

Science has been alright, it isn't my favorite subject of all time but it's been OK. I liked my 8th grade science class.

<u>Jessica (S8)</u>

I believe science was rather easy before coming to high school. It's gradually become harder. In middle school, the things we learned were rather simple and we mostly did labs. In high school it became harder because we've had to do work more with memorizing notes and understanding concept of biology instead of just doing labs.

Appendix E

Student Generated Questions

1. Why is the not-sanitized area of River City affecting the residents?
2. How the mosquitoes from the town dump are affecting residents of River City?
3. How the wealthy place in town is less likely to be affected by the mosquitoes?
4. Why is the problem getting worse day by day?
5. What correlation do algae and iron have in order to make atmospheric carbon?
6. What happened to the algae when too much iron was added to the ocean?
7. In what ways did the diatoms in the ocean survive from the krill?
8. How can algae bloom 100 meters deep in the ocean?
9. How can carbon be stored in the bottom of the ocean with seafloor sediments?
10. Do you think large-scale ocean iron dumping is good or bad?
11. What would happen if a large scale ocean iron dumping was done without further research and experiment?
12. Do you agree or disagree with the statement “The findings contribute to science’s understanding of the global carbon cycle and has implications for potential ways of mitigating rising levels of carbon dioxide, which contributes to climate change.” What evidence is there to support your answer?
13. What are some similarities and differences in dumping iron into algal blooms and phytoplankton blooms?
14. What are the objectives of Dr. Smetacek?
15. Why are carbohydrates important for your body?
16. How do I graph quadratic function?

17. How do you distinguish between a habitat and a niche?
18. Where is the DNA kept in a cell and where is RNA transcribed?
19. What is the main energy source of a cell?
20. What is the name of the site where proteins, ribosomes are constructed?
21. What are all the cells in the human body and what are their functions?
22. What are some analogies based on the cells in the human body and how they function with similar functioning things in the world?
23. What are some similarities between plant cells and animal cells? What are some differences?
24. What are some problems that arise in cells and how does this effect your body?
25. How would you improve your muscular endurance to become a better football player?
26. Are humans causing or contributing to global warming?
27. What role does the ocean paly in global warming?

Appendix F

A Sample Google Docs Posting

<table>
<tr><td colspan="2">Problem of the Day
Example: Why did so many people get sick in River City?</td></tr>
<tr><td rowspan="3">Initial ideas</td><td>1.</td></tr>
<tr><td>2.</td></tr>
<tr><td>3.</td></tr>
<tr><td colspan="2"></td></tr>
<tr><td rowspan="3">Questions</td><td>1.</td></tr>
<tr><td>2.</td></tr>
<tr><td>3.</td></tr>
<tr><td colspan="2">Gathered New Information Based on the Posed Questions</td></tr>
<tr><td colspan="2"></td></tr>
<tr><td colspan="2"></td></tr>
<tr><td>Arguments/ Disagreements among the Group Members</td><td>1. What do you think of …… and why?</td></tr>
<tr><td></td><td></td></tr>
<tr><td colspan="2"></td></tr>
<tr><td colspan="2"></td></tr>
<tr><td colspan="2"></td></tr>
</table>

Appendix G

A Sample Daily Email

River City Mailer <rivercity5@activeworlds.com to me May 31, 2016

Following is a list of notebook entries for all of your students in all of your class that used the online notebook in River City yesterday, by class, then name, then time. You will only receive this e-mail if any students made notebook entries on the prior day.

Please direct questions or comments regarding this e-mail to rivercity.support@activeworlds.com.

Please also note that all times are Eastern Time (ET). Thank you for your participation.

##

Class: Aghekyan

##

Tom at 8:03 pm (January 1879): Children's Cribs -- The pediatric ward is less full today, but Erica is here with her bad fever. Her mother is so sad and says that next summer she won't allow her to play near the bog at all!

Tom at 8:09 pm (January 1879): "Whitewings" - street sweepers -- These men spread sand in the warmer months to help cover up the horse manure. Now, why would they spread sand?

Tom at 8:09 pm (January 1879): Street looking toward factory -- There were a lot of workers from this factory that were sick 2 months ago. Most of them have recuperated and are back at work!

Tom at 8:11 pm (January 1879): Erica Loskill -- Erica says that she doesn't feel well and that her mother told her not to play in the bog during the summer time anymore. She gets hot and then gets cold.

Tom at 8:13 pm (January 1879): Observation and Inference -- If you see something happen then you observe it or you make an observation.

If you don't see something happen but you have some information, you could make an inference about what will happen next.

Tom at 8:40 pm (January 1879): Erica & The Bog -- Erica states that she played in the local bog during the summer. The Hospital Admissions Chart states that she visited on January 3rd and was experiencing an intense episode of fever and violent chills. Since Erica is living in tenement 1, I'm expecting many more people to get sick and have similar symptoms to what Erica had.

Tom at 8:45 pm (January 1879): Men and dog in front of boarding school or home -- The wealthier boys who are at boarding school miss their friends in River City. They heard that in the fall some of their friends were sick with stomachaches, but none of these boys got sick at their school which is several hours from River City.

Tom at 8:49 pm (January 1879): Wealthier Boys -- So far from the information that has been gathered, no wealthy child or adult has been sick or started showing signs of an illness. This is probably due to the fact that they live in a much better environment.

Tom at 8:58 pm (January 1879): Definition of Symptom -- Something that a living thing shows or indicates that a disease is present.

Tom at 9:00 pm (January 1879): Disease and Symptom -- When you have a disease, you experience

symptoms. For example you may have the flu, a disease, and one of your symptoms might be a fever.
Tom at 9:01 pm (January 1879): Day 2 Hospital Chart -- January Chat

Chapter 5. Discussion

The present set of studies aimed at investigating high school students' science identities, motivation and environmental attitudes in two different settings. First, it examined these constructs "as is" using a large and diverse group of students across many schools districts. Next, it looked how students' science identities changed and evolved during a collaborative online project called the River City. This second setting involved a small set of students. As a precursor to these investigations, a convenient, online survey instrument called SIEVA was designed and validated.

The multi-step, multi-study nature of this research necessitated the use of both quantitative and qualitative research methods. The design and validation of the SIEVEA instrument was the objective of the first study with three sub-studies. These studies used quantitative methods for exploring the instrument's factor structure and doing multiple reliability and validity tests. The second study also used quantitative methods since its goal was to analyze the collected large data set and discover significant relationships and dependencies between various data elements. The last study was unique. It had a small number of participants who were engaged in a web-based, video game-like learning environment. Because of the small number of participants, this study relied on qualitative research methods.

In this chapter, the results of all three studies will be discussed and summarized. Further, implications of this research for researchers and educators will be explained. Finally, limitations of the current research will be listed and future research directions will be outlined.

The SIEVEA Instrument

The SIEVEA survey instrument was essential for this study. It allowed to collect a large amount of data that was subsequently used to explore the research constructs as well as students' interest in science subjects along with student demographic data.

The design of the survey items was based on existing literature regarding science identities (Carlone & Johnson, 2007), expectation of success (Eccles et al., 1998), values of science (Eccles & Wigfield, 1995) and environmental attitudes (Dunlap & Van Liere, 1978; Dunlap, Van Liere, Mertig, & Jones, 2000). While designing this survey, it was important to rely on existing research and survey instruments to ensure that the survey was a valid instrument for measuring the constructs and its results could reliably be used to do further data analysis. Thus, while designing the survey's items, an extensive review of existing survey instruments and prior research in questionnaire design and constructions was conducted. This research considered both motivational/interest surveys (Fortus & Vedder-Weiss, 2014; Glynn, Brickman, Armstrong, & Taasoobshirazi, 2011; Glynn & Koballa, 2006; Plantae, O'Keefe, & Theoret, 2012; Swarat, Ortony, & Revelle, 2012) and environmental attitude/behavior surveys (Blatt, 2014; Bradley, Waliczek, & Zajicek, 1999; Eilam & Trop, 2012; Johnson & Manoli, 2011; Zecha, 2010).

Even though the design of the survey relied on existing instruments, necessary modifications were made in order to satisfy this study's objectives. Reviewing and piloting the survey provided additional steps of survey's refinement and improvement before it was used by a large group of students.

A considerable part of this research was dedicated to evaluating the SIEVEA instrument's validity and reliability since these two elements are critical for assessing survey instruments (Tavakol & Dennick, 2011). During the initial stage of the study, exploratory factor analysis were conducted on the SIEVEA survey's data. The first objective was to discover the factor structure of the instrument using the collected large data set. Instead of using the conventional, but not very reliable, eigenvalue-greater-than-one factor extraction method (Kaiser, 1960), this study used Parallel Analysis method developed by Horn (1965) since it is known to produce more accurate and reliable results than other factor extraction methods (Glorfeld, 1995). After applying Parallel Analysis to the results of the exploratory factor analysis, three candidate models were produced. The candidate factor structures mapped well to the a priori structure of the research constructs as suggested by the design of the survey. However, these candidate models suggested alternate ways of grouping the survey's items with varying degrees of fidelity to the designed, a priori structure. Therefore, additional analysis were needed in order to pick the final model.

First, various close-fit indexes like absolute close-fit indexes and incremental close-fit indexes (Hu & Bentler, 1999) were computed for all models. The indexes provided reliable measures for evaluating each model's fit to data. Next, all models were analyzed on how well they aligned with the research constructs. Last, the models were assessed based on their factor loadings. These analyses revealed that the three-factor model was the best model out of the three candidate models. Therefore, it was selected as the final model. Additionally, the Cronbach's alphas were measured for each group of items representing the research constructs. All alpha values were within the range of "good" values as defined by DeVellis (2003) indicating a high degree of survey's reliability.

Next, the partial-confirmatory and confirmatory factor analyses were conducted on the three-factor model in order to confirm this model's fit to data and perform various validation tests on it. The results of these analyses allowed assessing the three-factor model across several dimensions. First, its factor structure and fit to data were confirmed by conducting restricted factor analysis and computing various absolute and incremental close-fit indexes. The values of all these indexes were within their acceptable ranges, thus, confirming the model's fit to data. In addition, the factor loadings produced by confirmatory factor analysis were consistent with the loadings produced by exploratory factor analysis. Next, the three-factor model was evaluated for measurement invariance across different groups to ensure the instrument's data can be used for consistent comparison of differences between these groups (Schoot, Lugtig, & Hox, 2012). Last, stringent reliability and validity tests were applied to the model. These tests provided a strong evidence of the instrument's reliability and convergent and discriminant validity and confirmed that there were no major validity issues with the SIEVEA.

Lastly, the polytomous Rasch analysis (Andrich, 1988; Rasch, 1960) was used to evaluate the psychometric properties of the three-factor model and convert the ordinal scores of the survey's data to interval scale prior to conducting parametric statistical tests like ANOVA. The Rasch analysis reinforced the case of the SIEVEA instrument's reliability and validity. Both item and person fits were evaluated. In general, this analysis established a good fit for the three-factor model instrument, although some areas of future improvements were also revealed. The discovered issues were primarily with item and person misfits and outlier responses.

To summarize, this study followed a well-established and rigorous path for designing, testing and validating the SIEVEA survey instrument. Its reliability and validity were checked and confirmed using appropriate methods and procedures. While this process was challenging and laborious, it was necessary in order to establish the SIEVEA instrument's suitability and reliability for this and future research.

High School Students' Identities, Motivation, Interest and Environmental Attitudes

The usefulness and power of the SIEVEA survey instrument was apparent while investigating the current state of the research constructs for a large group of students. Indeed, once the districts' permission to participate in the survey was secured, it did not take much effort to set up a link and make the survey available for students. This easy of setup and use made it possible to collect a large sample of data: 3,454 student participants from 13 urban, suburban and private schools in three different states: New Jersey, Pennsylvania and Connecticut. Likewise, the collected responses provided a rich, meaningful data for conducting numerous quantitative analyses and uncovering significant findings related to both the research constructs and their relations to students' science preferences and student demographic data like gender and school type.

It turned out that students' science identities and motivation in science, values of science and environmental attitudes were significantly affected by their science subject preferences. The students' who preferred biology, chemistry and physics had strong science identities and motivation in science. Likewise, chemistry and physics preference led to a stronger appreciation of usefulness of science, whereas environmental science preference was a good predictor of stronger environmental attitudes. These results should not be surprising. Indeed, favoring "premium" science subjects, like physics or chemistry, should relate to strong science identity and motivation to learn science. Also, it is expected that these students will highly value science. Likewise, students with strong environmental attitudes should be interested in learning environmental science, which covers lots of contemporary environmental issues and topics related to the preservation of the environment and sustainability. Indeed, this result was similar to the findings of Bradley et al. (1999) research, which showed that students who were taught a course related to the environment developed positive environmental attitudes.

According to Matsunaga (2010), most phenomena that are studied in social sciences are interrelated. This study provided another proof for Matsunaga's statement by establishing that students' science identities and motivation in science, values of science and attitudes toward the environment were correlated with statistically significant correlation coefficients. Moreover, the discovery of correlations between the motivational and environmental attitude constructs was in line with the findings of De Groot and Steg (2010) who observed that pro-environmental behaviors could be explained by motivation using the Deci and Ryan (2000) Self-Determination Theory. Furthermore, both the values and self-determination motivation are related to pro-environmental behaviors, including recycling, using environmentally friendly products, and lessening the ecological footprint (De Groot & Steg, 2010).

The gender-related dissimilarities was another area of significant discovery. The analyses of the SIEVEA data showed statistically significant differences in science identities and motivation of males and females and in how they value science. It turned out that males had stronger science identities and were more motivated to learn science than females. Interestingly, these differences were most pronounced among students who favored physical sciences. These gender-related differences were in agreement with the findings of similar studies that looked into how students' attitudes toward science and their achievement in science varied by gender and other variables like race, home environment, parents' education and amount of homework (Schibeci & Riley, 1986; Weinburgh, 1995). According to the results of these studies, males have more positive attitudes toward science than females (Weinburgh, 1995). However, the gender differences discovered in the current study were different from those in the Stark and Gray (1999) study. Since the current study was done with high school students, whereas Stark and Gray worked with elementary school students, it appears that as the students advance from elementary school to middle or high school, their motivation in science learning undergoes a significant transformation. This hypothesis should be explored using a longitudinal study with the same set of participants throughout their elementary, middle and high school years.

The analysis of the SIEVEA data related to student's attitudes toward the environment showed that difference between males' and females' environmental attitudes were not statistically significant. This outcome is remarkable and could be the result of growing public environmental awareness and/or the increased emphasis on environmental preservation and environmental sustainability in school curriculums.

Other gender-related differences involved students' science subject preferences. It turned out that there were statistically significant differences between females' and males' science subject preferences. Namely, males favored chemistry over other science subjects, whereas females had a string preference for biology. This findings were in line with the results of Schibeci's (1984) study, which found that gender differences varied based on the science subject: girls had more positive attitudes toward biology, whereas boys liked physics and chemistry.

Since the survey participants were from both urban and suburban schools, it became possible to conduct differential data analyses between these two school types. The results were quite interesting. First, it turned out that urban and suburban students had different science subject preferences. Although biology and chemistry were popular with both urban and suburban students, urban students showed a much stronger preference for forensics than suburban students. Also, these subject preference differences were statistically significant.

Next, a statistically significant difference was discovered in urban and suburban students' attitudes toward the environment. Specifically, the data analysis showed that urban students had better environmental attitudes than suburban students. This result was not surprising given Stern, Powell & Ardoin's (2011) discoveries regarding urban students' sense of environmental responsibility. Indeed, that study found that urban students were very receptive to views that advocated environmentally responsible policies. Last but not least, there were no statistically significant differences in urban and suburban students' science identities and motivation and in how they value science. This result was somewhat unexpected.

The River City: A Case of Identity Development

According to Lave & Wenger's (1991) sociocultural learning theory, learning shifts from knowledge development to identity development. Rubin (2007) argued that in the figured world of learning, everyday activities should validate students as learners: good quality classroom discourse, encouraging student participation, and stressing the importance of students' ideas help in shaping learning identities. The overall goal of the River City study was to investigate how computer-based learning affects students' science identities, with the main focus on the changes in the urban youth's science identities.

Research in the past found that although many students understood major concepts in the Environmental Science course and scored well on their tests, they were not able to answer questions based on the real-life application of the content that they had "mastered" (Lord, 1999). Hmelo-Silver, Duncan & Chinn (2007) argued that although Problem Based Learning (PBL) and Inquiry Learning (IL) have different origins, both pedagogical practices involve authentic learning and sometimes are indistinguishable from each other. The learning environment of the River City study allowed students to work on real-world problems, which reflect genuine professional activities. These types of problems are frequently absent in academic curricula (Chinn & Malhotra, 2002). From a research standpoint, the graphical representations are essential for comprehending complex spatial structures, whereby simulating deeper understanding of concepts (Naaz, Chariker, & Pani, 2014).

Similarly, a big proponent of digital learning and game literacy, Gee (2007) purported that the game's interactive nature motivates students while making their learning meaningful. Because of the River City's game-like and graphical interface, students were able to grasp complex situations and gained thoughtful insights regarding the mysterious sickness that was ravaging River City. As one of the students, Samad, mentioned, "River City has helped me gain a better understanding of how science works, because sometimes you don't know what is causing the problem and it can lead to frustration and confusion."

According to Moje et al. (2007), science identities are not stagnant, but rather they evolve. Therefore, it was expected that students' identities would shift towards more scientific identities as a result of their participation in the River City project. The qualitative analysis of data generated by various instruments used during the project showed certain changes in students' science identities. First, students' letters to the Mayor of River City provided an evidence of change in students' science identities during the River City project. Both competence and performance aspects of science identity, as defined by Carlone and Johnson (2007), were present in students' initial letters to the mayor. As the project progressed, the students' competence and performance traits improved as evidenced by their final letters to the mayor.

Next, the analysis of the SIEVEA instrument's data supported the results of students' letters' examination: it also showed changes in students' science identities. For example, there was a positive shift in answers of 5 students out of total 8 when answering the following survey item: "My friends and family recognize me as a scientist." The remaining 3 students' answered showed no change. Lastly, the SIQ questionnaire answers provided more evidence of change in students' science identities during the River City project. Most students reported increased confidence in their abilities in learning science while answering SIQ questions "Are you confident in your abilities in science classes? " and "Do you think you will feel more confident in your abilities in science classes after your experiences in the River City?" All these results indicated that students' science identities underwent changes while they worked on real-world, environmental problems of the River City project. However, the results were not definitive. Indeed, due to the small sample size, the short duration of the project and measurement errors, the correctness of these results cannot be confidently validated without an additional, longitudinal research involving more participants.

Wang et al. (2014) validated that the exploitation of technology assisted in science learning in various science domains and topics related to science. In addition, Wang and his colleagues argued that the lecture-format, used in teacher-centered teaching, is less effective and requires low cognitive skills, in comparison with student-centered learning, which can promote critical thinking skills in students. Wang et al. (2014) findings added to the prior research that claimed that inquiry-based science curricula helped urban boys develop skills that are necessary to succeed on standardized tests (Geier, Blumenfeld, Marx, Krajcik, Fishman, Soloway and Clay-Chambers, 2008). This success was largely due to the use of technology, peer collaboration and inquiry; important learning strategies that often lack in traditional instruction. The River City's environment provided participating students with lots of opportunities to use technology for science learning. Additionally, by engaging students in student-centered, problem-based learning, the River City project helped them in practicing and enhancing 21st century skills. As a result, there was improvement in students' problem-solving and critical thinking skills.

The overall results of the River City and student experiences during the project demonstrated the usefulness of technology-based, student-centered teaching strategies. Without a doubt, the River City project was a positive learning experience for students and facilitated the development of their 21st century skills.

Implications for Research Practice

As this study revealed, the SIEVEA survey is a simple and usefulness instrument for measuring students'' science learning related constructs like science identities, motivation and environmental attitudes. The web-based nature of the survey makes it a convenient tool for collecting a large amount of data with little effort. Indeed, this research is a testament to this statement. The researcher was able to set up multiple instances of the survey to collect lots of data from many schools across several states.

Additionally, several valuable and convenient features of the SIEVEA instrument help in assuring a high level of student participation and completion of the survey. First, the survey can be taken either from home or in the classroom. Next, the survey's questions are age appropriate for high school students as established by the Lexile framework. Then, since the survey is short, it does not take much time to complete the survey: about 10 minutes.

All these features make the SIEVEA survey a good tool for researchers who study students' science identities, expectations of success in science, values of science and environmental attitudes. It allows for performing a rich, quantitative research using multiple statistical tests and for exploring the impact of students' gender, school type, and favorite science subject on above-mentioned constructs. The researchers can use the survey as many times as needed during their research in orer to collect data related to science learning and environmental attitudes and measure the usefulness of their researched methods on aforementioned constructs. Moreover, since the survey is online, it is very convenient to extract participants' responses for further quantitative data analyzes.

Implications for Educational Practice

The SIEVEA instrument can also be utilized by the school administration and teachers to survey total student population's science and environmental attitudes. They can use the data generated by the survey to discover areas of concern in their schools and make necessary changes in their curriculum, instruction and teaching strategies. Identifying problems with students' motivation in science learning is a good example of how the school personnel can benefit from SIEVEA's data. Once the areas of concern are discovered, specific recommendations can be made for fixing the problems and improving students' motivation.

The River City project demonstrated the power of technology-based, multi-user virtual learning environments. During this project students worked on the real-world problems with very little supervision. While exploring health and environmental problems plaguing River City, students acted as a scientist and learned how to use technology, collect and analyze data, form hypothesis, and make scientifically sound conclusions regarding the observed phenomena. This project taught students 21st century skills and helped them to become independent learners. Therefore, it is highly recommended that the educators use this type of learning environments in their teaching practice and integrate them with their existing teaching strategies.

Limitations

The SIEVEA instrument is a valuable tool for collecting and analyzing data regarding student's science identities, motivational and environmental attitudes. However, several limitations with the collected data and instrument should be noted. First, even though data was collected from both urban and suburban schools, no other demographic data was collected except for gender due to the anonymous nature of the study. This helped in convincing more school districts to participate. However, this made it impossible to collect and use other important demographic data of the participants like students' socio-economic status, race, ethnicity, etc. Because of this, no analysis could be done regarding these demographic elements.

Another limitation of the study was that students' grade level information was not collected. Therefore, the study's results gave a combined picture of all high school grade levels. It will be beneficial to collect grade level data in future studies so results of the study can be cross-validated across different high school grades. Next, the science values factor had only two items in the survey. This fact limited the usability of this factor. Additionally, the Rasch analysis does not work well for constructs with less than three items. Consequently, the results of parametric statistical tests pertaining to this factor should be used with a grain of salt. This shortcoming of the SIEVEA survey instrument can be improved by adding more items related to the science value construct.

The majority of survey's use only positively worded items (Usher & Pajares, 2009). The SIEVEA was not an exception; it included only positively worded items. Some researchers recommended using both positively and negatively worded items in the same survey since it can improve the survey's ability to differentiate between extreme and moderate responses and reduce the ceiling and basement effects (Spector, Van Katwyk, Brannick, & Chen, 1997). Therefore, future revisions of the SIEVEA should include changing item wording in order to have several negatively worded items along with positively worded ones.

Lastly, the River City project showed certain positive shifts in students' science identities. However, because the River City project was short-term (it took only two weeks), its results were not conclusive and cannot be used to predict how students' science identities will evolve and persist over longer time intervals.

Future Directions

Expanding the application of the SIEVEA survey across time and geography is a worthwhile future exercise. One such an exercise will be to understand how high school students' science identities, motivation in science and environmental attitudes evolve during their high school years. Since the SIEVEA survey was designed for high school students, it is suitable for conducting this type of longitudinal studies. Indeed, the researchers can use the SIEVEA instrument with the same students during their 9th to 12th grades. After collecting data for four consecutive years, they will be able to examine how high school students' science identities, motivation in science and environmental attitudes change while they advance from one grade to the next grade in high school. This kind of research will be very useful because it may help in finding either positive or negative transitional periods during the high school years. Then, these transitional periods can be further scrutinized in order to find their underlying reasons and apply corrective interventions if necessary.

Another example of recommended longitudinal study is to conduct the SIEVEA survey twice a year: at the beginning and at the end of the year. This will help in determining how students' science identities, motivation and environmental attitudes change as they complete the academic year. It will also be useful to expand the geography of SIEVEA's data collection. This study's data was collected from various school districts in three states of the U.S.A.: New Jersey, Pennsylvania and Connecticut. Since the survey is web-based, it is feasible to collect data from school districts in other states, as well as from schools located in other countries. By expanding the geography of SIEVEA's data collection, more diverse data can be collected, leading to improved external validity of future studies that will analyze the collected data. For example, de Groot & Steg (2007) suggested that the environmental attitudes of high school students could deviate across different countries. Therefore, expanding the SIEVEA's data collection geography to other countries will provide means for validating de Groot & Steg's (2007) results.

Critical thinking skills are becoming increasingly important for 21st century workforce. This necessitates the use of such instructional approaches in schools that help students in fostering these skills. Students can enhance their critical thinking skills by formulating alternative explanations (Duschl, Schweingruber, & Shouse, 2007; Russell, Lucas & McRobbie, 2004) while reflecting on their own and their peer's thinking (McNeil & Pimentel, 2010). Therefore, it was expected that the River City project would enhance and promote students' critical thinking skills, since students were engaged in tasks that required them to evaluate unfamiliar situations, solve the problems and make decisions. Dianna Kuhn (2007) stated, "A steady diet of 'worked problems' cannot possibly prepare today's students for what they will face in the 21st century world." Engaging students in ill-structured problem-solving PBL format environment of the River City could help them in practicing and enhancing many skills that are essential for 21st century citizens. Therefore, it is recommended to research how the River City project can influence high school students' critical thinking and problem-solving skills.

In problem-based learning environments like the River City, students can apply their acquired knowledge to new situations (Hmelo & Lin, 2000). Hence, it was expected that some level of knowledge transfer would take place during the River City project. However, the current study did not investigate the transfer of knowledge during this project. Therefore, it is recommended to conduct a future study exploring if and how the knowledge transfer happens during learning activities in virtual worlds similar to the River City.

Conclusions

Both national (NAEP, 2015) and international (PISA, 2012) student assessments programs show that most U.S. students lack scientific literacy skills even though these skills became increasingly important in 21st century. Therefore, it is imperative to understand what motivates students to learn science and how to maintain their motivation as they progress through school years. In order to understand what motivates students to learn science, it is useful to examine students' reasons of science learning and their beliefs that influence this learning (Glynn et al., 2008). Many studies of motivation indicated that students, who hold strong beliefs and expectations about themselves, are more likely to engage in tasks and persevere in doing them despite any challenges (Eccles et al., 1998). In addition, students with strong science identities are more likely to participate in science classes and succeed (Sfard & Prusak, 2005).

This study embarked on a journey to explore high school students' science identities, motivation and environmental attitudes. While doing this, a new instrument, called SIEVEA, was developed for measuring these construct. Then, the instrument was successfully used in subsequent studies where aforementioned constructs were measured and analyzed in two different contexts: 1) for a large group of students from multiple schools in three different states and 2) for a small group of students from an urban school participating in the learning project within the multi-user virtual learning environment of the River City.

The study produced many significant results that are of interest to both the research and educational communities concerned with student science learning. The analysis of data helped to uncover many expected and some unexpected results which will help to comprehend how students' science identities, motivation and environmental attitudes vary by gender, interest, school type. Additionally, students' science subject preferences were examined and rationalized. Finally, the River City project helped in understanding how urban students' science identities evolve when they work on the real-life problem in virtual learning environments.

The results of this study have both theoretical and practical values. They add to the existing knowledge of science identities, motivation and environmental attitudes gained by prior research. They also provide useful insights to educational practitioners for developing instructional and teaching strategies, which can promote stronger student science identities, increase students' motivation and enhance their environmental attitudes.

References

Andrich, D. (1988). *Rasch Models for Measurement*: Thousand Oaks, CA: Sage.

Blatt, E. (2014). Uncovering students' environmental identity: An exploration of activities in an environmental science course. *The Journal of Environmental Education*, *45*(3), 194-216.

Bradley, J. C., Waliczek, T. M., & Zajicek, J. M. (1999). Relationship between environmental knowledge and environmental attitude of high school students. *The Journal of Environmental Education, 30*(3), 17-21.

Carlone, H. B., & Johnson, A. (2007). Understanding the science experiences of successful women of color: Science identity as an analytic lens. *Journal of Research in Science Teaching, 44*(8), 1187-1218.

Chinn, C. A., & Malhotra, B. A. (2002). Epistemologically authentic inquiry in schools: A theoretical framework for evaluating inquiry tasks. *Science Education*, *86*(2), 175-218.

De Groot, J. I., & Steg, L. (2007). Value orientations and environmental beliefs in five countries validity of an instrument to measure egoistic, altruistic and biospheric value orientations. *Journal of Cross-Cultural Psychology, 38*(3), 318-332.

Deci, E. L., & Ryan, R. M. (2000). The" what" and" why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry, 11*(4), 227-268.

DeVellis, R. (2003). *Scale development: Theory and applications.* Thousand Oaks, CA: Sage.

Dunlap, R. E., & Van Liere, K. D. (1978). The "new environmental paradigm". *The Journal of Environmental Education, 9*(4), 10-19.

Dunlap, R. E., Van Liere, K. D., Mertig, A. G., & Jones, R. E. (2000). New trends in measuring environmental attitudes: measuring endorsement of the new ecological paradigm: a revised NEP scale. *Journal of Social Issues*, *56*(3), 425-442.

Duschl, R., Schweingruber, H., & Shouse, A., (Eds.), (2007). *Taking Science to School: Learning and Teaching Science in Grades K-8.* Washington, DC : National Academies Press.

Eccles, J. S., & Wigfield, A. (1995). In the mind of the actor: The structure of adolescents' achievement task values and expectancy-related beliefs. *Personality and Social Psychology Bulletin, 21*(3), 215-225.

Eccles, J. S., Wigfield, A., & Schiefele, U. (1998). Motivation to succeed. In W. Damon (Series Ed.) & N. Eisenberg (Vol. Ed.), *Handbook of Child Psychology: Vol. 3.*

Social, emotional and personality development (5th ed., pp. 1017–1094). New York, NY: Wiley.

Eilam, E., & Trop, T. (2012). Environmental attitudes and environmental behavior—Which is the horse and which is the cart? *Sustainability*, *4*(9), 2210-2246.

Fortus, D., & Vedder-Weiss, D. (2014). Measuring students' continuing motivation for science learning. *Journal of Research in Science Teaching, 51*(4), 497-522.

Gee, J. P. (2007). *What video games have to teach us about learning and literacy*, 2nd Edition. New York: Palgrave Macmillan.

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

Glynn, S. M., Brickman, P., Armstrong, N., & Taasoobshirazi, G. (2011). Science motivation questionnaire II: Validation with science majors and nonscience majors. *Journal of Research in Science Teaching, 48*(10), 1159-1176.

Glynn, S. M., & Koballa, T. R. (2006). Motivation to learn in college science. In J. J. Mintzes & W. H. Leonard (Eds.), *Handbook of college science teaching* (pp. 25-32). Arlington, VA: National Science Teachers Association Press.

Glorfeld, L.W. (1995). An improvement on Horn's parallel analysis methodology for selecting the correct number of factors to retain. *Educational and Psychological Measurement, 55*, 377-393.

Hmelo, C. E., & Lin, X. (2000). Becoming self-directed learners: Strategy development in problem-based learning. In D. Evensen & C. E. Hmelo (Eds.), *Problem-based learning: A research perspective on learning interactions* (pp. 227-250). Mahwah, NJ: Erlbaum.

Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist*, *42*(2), 99-107.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*, 179-185.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1-55.

Johnson, B., & Manoli, C. C. (2011). The 2-MEV scale in the United States: A measure of children's environmental attitudes based on the theory of ecological attitude. *The Journal of Environmental Education*, *42*(2), 84-97.

Kaiser, H.F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement, 20,* 141-151.

Kuhn, D. (2007). Is direct instruction an answer to the right question? *Educational Psychologist, 42*(2), 109-113.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. New York: Cambridge University Press.

Lord, R. T. (1999). A comparison between traditional and constructivist teaching in environmental science. *The Journal of Environmental Education*, *30*(3), 22-28.

Matsunaga, M. (2010). How to factor-analyze your data right: Do's, don'ts, and how-to's. *International Journal of Psychological Research*, *3*(1), 97-110.

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

Naaz, F., Chariker, J. H., & Pani, J. R. (2014). Computer-based learning: Graphical integration of whole and sectional neuroanatomy improves long-term retention. *Cognition and Instruction*, *32*(1), 44-64.

PISA. (2012). *Results in Focus: What 15-year-olds know and what they can do with what they know.* Retrieved from http://www.oecd.org/pisa/keyfindings/pisa-2012-results-overview.pdf

Plante, I., O'Keefe, P. A., & Théorêt, M. (2013). The relation between achievement goal and expectancy-value theories in predicting achievement-related outcomes: A test of four theoretical conceptions. *Motivation and Emotion*, *37*(1), 65-78.

Rasch, G. (1960). *Probabilistic models for some intelligence and achievement tests.* Copenhagen: Danish Institute for Educational Research.

Rubin, B. C. (2007). Learner identity amid figured worlds: Constructing (in) competence at an urban high school. *The Urban Review*, *39*(2), 217-249.

Russell, D. W., Lucas, K. B., & McRobbie, C. J. (2004). Role of the microcomputer-based laboratory display in supporting the construction of new understandings in thermal physics. *Journal of Research in Science Teaching, 41*(2), 165-185.

Schibeci, R. A. (1984). Attitudes to science: An update. *Studies in Science Education, 11*, 26-59.

Schibeci, R. A., & Riley, J. P. (1986). Influence of students' background and perceptions on science attitudes and achievement. *Journal of Research in Science Teaching, 23*(3), 177-187.

Sfard, A., & Prusak, A. (2005). Telling identities: In search of an analytic tool for investigating learning as a culturally shaped activity. *Educational Researcher*, *34*(4), 14-22.

Spector, P. E., Van Katwyk, P. T., Brannick, M. T., & Chen, P. Y. (1997). When two factors don't reflect two constructs: How item characteristics can produce artifactual factors. *Journal of Management*, *23*(5), 659-677.

Stark, R., & Gray, D. (1999). Gender preferences in learning science. *International Journal of Science Education, 21*(6), 633-643.

Stern, J. M., Powell, B. R., & Ardoin, M. N. (2011). Evaluating a constructivist and culturally responsive approach to environmental education for diverse audiences. *The Journal of Environmental Education*, *42*(2), 109-122.

Swarat, S., Ortony, A., & Revelle, W. (2012). Activity matters: Understanding student interest in school science. *Journal of Research in Science Teaching*, *49*(4), 515-537.

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, *2*, 53-55.

Usher, E. L., & Pajares, F. (2009). Sources of self-efficacy in mathematics: A validation study. *Contemporary Educational Psychology*, *34*(1), 89-101.

Van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology*, *9*(4), 486-492.

Wang, C. Y., Wu, H. K., Lee, S. W. Y., Hwang, F. K., Chang, H. Y., Wu, Y. T., & Tsai, C. C. (2014). A review of research on technology-assisted school science laboratories. *Educational Technology & Society*, *17*(2), 307-320.

Weinburgh, M. (1995). Gender differences in student attitudes toward science: A meta-analysis of the literature from 1970 to 1991. *Journal of Research in Science Teaching, 32*(4), 387-398.

Zecha, S. (2010). Environmental knowledge, attitudes and actions of Bavarian (southern Germany) and Asturian (northern Spain) adolescents. *International Research in Geographical and Environmental Education, 19*(3), 227-240.

Bibliography

Agarwal, N. K. (2011). Verifying survey items for construct validity: A two stage sorting procedure for questionnaire design in information behavior research. *Proceedings of the American Society for Information Science and Technology, 48*(1), 1-8.

Aghekyan, R. (2015). Questioning Patterns during Discussions in Collaborative Groups in Socioeconomically Diverse High Schools. *NERA Conference Proceedings 2014.* Paper 6. http://digitalcommons.uconn.edu/nera_2014/6

Ainley, M. D. (2004). *Some perspectives on interest in learning and classroom interaction.* Retrieved from http://www.aare.edu.au/98pap/ain98054.htm

Aldous, D. E. (2010). The people-plant-place paradigm. In J. Janick, G. R. Dixon, & L. Rallo (Eds.), *XXVIII international horticultural congress on science and horticulture for people (IHC2010): Colloquia and overview* (pp. 171-178). Retrieved from http://www.actahort.org/books/916/

Alwin, D. F., Cohen, R. L., & Newcomb, T. M. (1991). *Political attitudes over the life span: The Bennington women after fifty years.* Madison, WI: University of Wisconsin Press.

American Association for the Advancement of Science. (1989). *Science for all Americans*. New York, NY: Oxford University Press. Also available at http://www.project2061.org/publications/sfaa/online/sfaatoc.htm

American Association of University Women Educational Foundation. (1992). *How schools shortchange girls: Executive summary.* Washington, DC: Author.

Aminrad, Z., Zakariya, S. Z. B. S., Hadi, A. S., & Sakari, M. (2013). Relationship between awareness, knowledge and attitudes towards environmental education among secondary school students in Malaysia. *World Applied Sciences Journal, 22*(9), 1326-1333.

Andersen, L., & Ward, T. J. (2014). Expectancy-value models for the STEM persistence plans of ninth-grade, high-ability Students: A comparison between Black, Hispanic, and White students. *Science Education*, *98*(2), 216-242.

Anderson, J. R., Reder, L. M., & Simon, H. A. (1996). Situated learning and education. *Educational Researcher*, *25*(4), 5-11.

Anderson, L. W., Krathwohl, D. R., & Bloom, B. S. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives.* Boston, MA: Allyn & Bacon.

Andre, T., Whigham, M., Hendrickson, A., & Chambers, S. (1999). Competency beliefs, positive affect, and gender stereotypes of elementary students and their parents about science versus other school subjects. *Journal of Research in Science Teaching, 36*(6), 719-747.

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika, 43*(4), 561-573.

Andrich, D. (1988). *Rasch Models for Measurement*: Thousand Oaks, CA: Sage.

Andrich, D., Lyne, A., Sheridan, B., & Luo, G. (2011). *RUMM 2030.* Perth: RUMM Laboratory. Also available at http://www.rummlab.com.au/

Arbuckle, J. L. (2016). *Amos user's guide version 24*. Chicago, IL: SmallWaters.

Aschbacher, P. R., Li, E., & Roth, E. J. (2010). Is science me? High school students' identities, participation and aspirations in science, engineering, and medicine. *Journal of Research in Science Teaching, 47*(5), 564-582.

Atkinson, J. W. (1957). Motivational determinants of risk-taking behavior. *Psychological Review, 64*, 359-372.

Atkinson, J. W. (1964). *An introduction to motivation*. Princeton, NJ: Van Nostrand.

Bandalos, D. L. (1996). Confirmatory factor analysis. In J. Stevens (Ed.), *Applied multivariate statistics for the behavioral sciences* (3rd ed., pp. 389-420). Mahwah, NJ: LEA.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Macmillan.

Baram Tsabari, A., Sethi, R. J., Bry, L., & Yarden, A. (2006). Using questions sent to an Ask-A-Scientist site to identify children's interests in science. *Science Education, 90*(6), 1050-1072.

Barton, A. C., Kang, H., Tan, E., O'Neill, T. B., Bautista-Guerra, J., & Brecklin, C. (2013). Crafting a future in science tracing middle school girls' identity work over time and space. *American Educational Research Journal, 50*(1), 37-75.

Bathgate, M. E., Schunn, C. D., & Correnti, R. (2014). Children's motivation toward science across contexts, manner of interaction, and topic. *Science Education, 98*(2), 189-215.

Baumgartner, H., & Hombur, C. (1996). Applications of structural equation modeling in marketing and consumer research: A review. *International Journal of Research in Marketing, 13*, 139-161.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*, 238–246.

Bentler, P. M. (1995). *EQS structural equations program manual.* Encino, CA: Multivariate Software.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*, 588–606.

Biswas, G., Schwartz, D. L., Leelawong, K., Vye, N., & TAG-V. (2005). Learning by teaching: A new paradigm for educational software. *Applied Artificial Intelligence, 19*(3).

Blatt, E. (2014). Uncovering students' environmental identity: An exploration of activities in an environmental science course. *The Journal of Environmental Education*, *45*(3), 194-216.

Bloom, B. S., Engelhart, M. B., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals* (Handbook 1: Cognitive domain). New York: Longmans Green.

Bollen, K. A. (1989). A new incremental fit index for general structural equation models. *Sociological Methods & Research*, *17*(3), 303-316.

Bollen, K. A. (1989). *Structural equations with latent variables*. NY: Wiley.

Bollen, K. A., & Stine, R. A. (1992). Bootstrapping goodness-of-fit measures in structural equation models. *Sociological Methods & Research, 21*, 205–229.

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2^{nd} ed.). Mahwah, NJ: Erlbaum.

Boone, W. J., & Scantlebury, K. (2006). The role of Rasch analysis when conducting science education research utilizing multiple-choice tests. *Science Education*, *90*(2), 253-269.

Bradley, J. C., Waliczek, T. M., & Zajicek, J. M. (1999). Relationship between environmental knowledge and environmental attitude of high school students. *The Journal of Environmental Education, 30*(3), 17-21.

Bradley, M. E., Thom, L. R., Hayes, J., & Hay, C. (2008). Ask and you will receive: How question type influences quantity and quality of online discussions. *British Journal of Educational Technology*, *39*(5), 888-900.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press.

Brentani, E., & Golia, S. (2007). Unidimensionality in the Rasch model: How to detect and interpret. *Statistica, 67*(3), 253-261.

Brickhouse, N., Lowery, P., & Schultz, K. (2000). What kind of a girl does science? The construction of school science identities. *Journal of Research in Science Teaching, 37,* 441–458.

Brickhouse, N.W. (2000). Embodying science: A feminist perspective on learning. *Journal of Research in Science Teaching*, 38, 282–295

Brill, G., & Yarden, A. (2003). Learning biology through research papers: A stimulus for question-asking by high-school students. *Cell Biology Education*, *2*(4), 266-274.

Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. New York, NY: Guilford Press.

Browne, M. W. (1990). *MUTMUM PC user's guide*. Columbus: Ohio State University, Department of Psychology.

Browne, M. W., & Cudek, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing Structural Equation Models* (pp. 136–162). Newbury Park, CA: Sage.

Bruner, J. (1991). The narrative construction of reality. *Critical Inquiry, 18*, 1-21.

Bryan, L. A., & Atwater, M. M. (2002). Teacher beliefs and cultural models: A challenge for science teacher preparation programs. *Science Education*, *86,* 821–839.

Burke, P. J. (1980). The self: Measurement requirements from an interactionist perspective. *Social Psychology Quarterly*, 18-29.

Burn, A., Buckingham, D., Parry, P., & Powell, M. (2010). Minding the gaps: Teachers' cultures, students' cultures. In D. E. Alvermann (Ed.), *Adolescents' online literacies: Connecting classrooms, digital media, and popular culture* (pp. 183-201). New York: Peter Lang.

Bybee, R. W., & Fuchs, B. (2006). Preparing the 21st century workforce: A new reform in science and technology education. *Journal of Research in Science Teaching*, *43*(4), 349-352.

Bybee, R., & McCrae, B. (2011). Scientific literacy and student attitudes: Perspectives from PISA 2006 science. *International Journal of Science Education, 33*(1), 7–26.

Byrne, B. M. (2013). *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. New York, NY: Routledge.

Byrne, K. A., Patrick, C. J., & Worthy, D. A. (2016). Striatal dopamine, externalizing proneness, and substance abuse effects on wanting and learning during reward-based decision making. *Clinical Psychological Science*, *4*(5), 760-774.

Campbell, D. T. & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56-81.

Cappelleri, J. C., Lundy, J. J., & Hays, R. D. (2014). Overview of classical test theory and item response theory for the quantitative assessment of items in developing patient-reported outcomes measures. *Clinical Therapeutics*, *36*(5), 648-662.

Carlone, H. B., & Johnson, A. (2007). Understanding the science experiences of successful women of color: Science identity as an analytic lens. *Journal of Research in Science Teaching*, *44*(8), 1187-1218.

Carr, D. (1998). The art of asking questions in the teaching of science. *School Science Review*, *79*, 47–50.

Chang, K. E., Chen, Y. L., Lin, H. Y., & Sung, Y. T. (2008). Effects of learning support in simulation-based physics learning. *Computers & Education*, *51*, 1486–1498.

Chi, M. T., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, *49*(4), 219-243.

Chiel, H. J., McManus, J. M., & Shaw, K. M. (2010). From biology to mathematical models and back: Teaching modeling to biology students, and biology to math and engineering students. *CBE-Life Sciences Education*, *9*(3), 248-265.

Child, D. (2006). *The essentials of factor analysis*. (3rd ed.). New York, NY: Continuum International Publishing Group.

Chin, C. (2006). Using self-questioning to promote pupils’ process skills thinking. *School Science Review*, *87*(321), 113–122.

Chin, C., & Chia, L. G. (2004). Problem-based learning: Using students' questions to drive knowledge construction. *Science Education*, *88*(5), 707-727.

Chin, C., & Kayalvizhi, G. (2005). What do pupils think of open science investigations? A study of Singaporean primary 6 pupils. *Educational Research*, *47*(1), 107–126.

Chin, C., & Osborne, J. (2008). Students' questions: a potential resource for teaching and learning science. *Studies in Science Education*, *44*(1), 1-39.

Chin, C., & Osborne, J. (2010). Supporting argumentation through students' questions: Case studies in science classrooms. *Journal of the Learning Sciences*, *19*(2), 230-284.

Chin, C., Brown, D. E., & Bruce, B. C. (2002). Student-generated questions: A meaningful aspect of learning in science. *International Journal of Science Education*, *24*(5), 521–549.

Chinn, C. A., & Malhotra, B. A. (2002). Epistemologically authentic inquiry in schools: A theoretical framework for evaluating inquiry tasks. *Science Education*, *86*(2), 175-218.

Cho, Y. H., Lee, J., & Jonassen, D. H. (2011). The role of tasks and epistemological beliefs in online peer questioning. *Computers & Education*, *56*(1), 112-126.

Choi, I., Land, S. M., & Turgeon, A. (2005). Scaffolding peer-questioning strategies to facilitate metacognition during online small group discussion. *Instructional Science*, *33*, 483–511.

Clarke, J., & Dede, C. (2009). Design for scalability: A case study of the River City curriculum. *Journal of Science Education and Technology*, *18*(4), 353-365.

Colucci-Gray, L., Camino, E., Barbiero, G., & Gray, D. (2006). From scientific literacy to sustainability literacy: An ecological framework for education. *Science Education*, *90*(2), 227-252.

Comer, J.P., Haynes, N.M., Joyner, E.T., & Ben-Avie, M. (1996). *Rallying the whole village: The Comer process for reforming education*. New York: Teachers College Press.

Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis*. Hillsdale, NJ: Erlbaum.

Costa, J., Caldeira, H., Gallastegui, J. R., & Otero, J. (2000). An analysis of question asking on scientific texts explaining natural phenomena. *Journal of Research in Science Teaching*, *37*(6), 602-614.

Costello, A. B., & Osborne, J. W. (2011). *Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis*. Pract Assess Res Eval 2005; 10. URL http://pareonline. net/getvn. asp, 10, 7.

Coyle, K. (2005). *Environmental literacy in America: What ten years of NEETF/Roper research and related studies say about environmental literacy in the U.S.*. Washington, DC: The National Environmental Education & Training Foundation. Retrieved from http://www.neefusa.org/pdf/ELR2005.pdf

Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches.* Thousand Oaks, CA: Sage.

Cuccio-Schirripa, S., & Steiner, H. E. (2000). Enhancement and analysis of science question level for middle school students. *Journal of Research in Science Teaching*, *37*(2), 210-224.

Cuevas, H. M., & Fiore, S. M. (2014). Enhancing learning outcomes in computer-based training via self-generated elaboration. *Instructional Science, 42*(6), 839-859.

Curtis, D. D., & Lawson, M. J. (2001). Exploring collaborative online learning. *Journal of Asynchronous Learning Networks, 5*(1), 21-34.

De Groot, J. I., & Steg, L. (2007). Value orientations and environmental beliefs in five countries validity of an instrument to measure egoistic, altruistic and biospheric value orientations. *Journal of Cross-Cultural Psychology, 38*(3), 318-332.

De Groot, J. I., & Steg, L. (2010). Relationships between value orientations, self-determined motivational types and pro-environmental behavioural intentions. *Journal of Environmental Psychology, 30*(4), 368-378.

de Koning-Veenstra, B., Steenbeek, H. W., van Dijk, M. W., & van Geert, P. L. (2014). Learning through movement: A comparison of learning fraction skills on a digital playful learning environment with a sedentary computer-task. *Learning and Individual Differences, 36*, 101-109.

Deci, E. L. (1992). The relation of interest to the motivation of behavior: A self-determination theory perspective. In K. A. Renninger, S. Hidi, & A. Krapp (Eds.), *The role of interest in learning and development* (pp. 43-70). Hillsdale, NJ: Erlbaum.

Deci, E. L., & Ryan, R. M. (2000). The" what" and" why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry, 11*(4), 227-268.

Dede, C., Ketelhut, D., & Ruess, K. (2003). *Designing for motivation and usability in a museum-based multi-user virtual environment*. Paper presented at the American Educational Research Association Conference, Chicago, IL

DeMars, C. E., & Erwin, T. D. (2004). Scoring neutral or unsure on an identity development instrument for higher education. *Research in Higher Education*, *45*(1), 83-95.

DeVellis, R. (2003). *Scale development: Theory and applications.* Thousand Oaks, CA: Sage.

DeVellis, R. F. (2012). *Scale development: Theory and applications* (3rd ed.). Los Angeles, CA: Sage.

Dewey, J. (1913). *Interest and effort in education.* Houghton Mifflin.

Dillon, J. T. (1988). The remedial status of student questioning. *Journal of Curriculum Studies*, *20*(3), 197-210.

Ding, L. (2012, February). Applying Rasch theory to evaluate the construct validity of brief electricity and magnetism assessment. In *2011 Physics Education Research Conference* (Vol. 1413, No. 1, pp. 175-178). AIP Publishing.

DiStefano, C., & Motl, R. W. (2006). Further investigating method effects associated with negatively worded items on self-report surveys. *Structural Equation Modeling*, *13*(3), 440-464.

Dresner, M., de Rivera, C., Fuccillo, K. K., & Chang, H. (2014). Improving higher-order thinking and knowledge retention in environmental science teaching. *BioScience*, *64*(1), 40-48.

Dunlap, R. E., & Van Liere, K. D. (1978). The "new environmental paradigm". *The Journal of Environmental Education, 9*(4), 10-19.

Dunlap, R. E., Van Liere, K. D., Mertig, A. G., & Jones, R. E. (2000). New trends in measuring environmental attitudes: measuring endorsement of the new ecological paradigm: a revised NEP scale. *Journal of Social Issues*, *56*(3), 425-442.

Durik, A. M., & Harackiewicz, J. M. (2007). Different strokes for different folks: How individual interest moderates the effects of situational factors on task interest. *Journal of Educational Psychology*, *99*(3), 597.

Duschl, R., & Grandy, R. E. (2013). Two views about explicitly teaching nature of science. *Science & Education*, *22*(9), 2109-2139.

Duschl, R., Schweingruber, H., & Shouse, A., (Eds.), (2007). *Taking Science to School: Learning and Teaching Science in Grades K-8.* Washington, DC : National Academies Press.

Eccles, J. S. (2005). Subjective task value and the Eccles et al. model of achievement-related choices. In A. J. Elliot & C. S. Dweck (Eds.), *Handbook of competence and motivation* (pp. 105-121). New York, NY: Guildford.

Eccles, J. S. (2009). Who am I and what am I going to do with my life? Personal and collective identities as motivators of action. *Educational Psychologist*, *44*, 78–89.

Eccles, J. S., & Wigfield, A. (1995). In the mind of the actor: The structure of adolescents' achievement task values and expectancy-related beliefs. *Personality and Social Psychology Bulletin*, *21*(3), 215-225.

Eccles, J. S., Wigfield, A., & Schiefele, U. (1998). Motivation to succeed. In W. Damon (Series Ed.) & N. Eisenberg (Vol. Ed.), *Handbook of Child Psychology: Vol. 3. Social, emotional and personality development* (5th ed., pp. 1017–1094). New York, NY: Wiley.

Eccles, J., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J. L., & Midgley, C. (1983). Expectancies, values, and academic behaviors. In J. T. Spence (Ed.), *Achievement and Achievement motives: Psychological and sociological approaches* (pp. 75-146). San Francisco, CA: W. H. Freeman.

Eckes, T. (2005). Examining rater effects in TestDaF writing and speaking performance assessments: A many-facet Rasch analysis. *Language Assessment Quarterly: An International Journal*, *2*(3), 197-221.

Edelson, D. C., & Joseph, D. M. (2004, June). The interest-driven learning design framework: motivating learning through usefulness. In *Proceedings of the 6th international conference on Learning sciences* (pp. 166-173). International Society of the Learning Sciences.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics, 7,* 1–26.

Eilam, E., & Trop, T. (2012). Environmental attitudes and environmental behavior—Which is the horse and which is the cart? *Sustainability*, *4*(9), 2210-2246.

Eom, K., Kim, H. S., Sherman, D. K., & Ishii, K. (2016). Cultural variability in the link between environmental concern and support for environmental action. *Psychological Science, 27*(10), 1331-1339.

Erdogan, M., Bahar, M., & Usak, M. (2012). Environmental education in high school 9th-12th biology course curricula started to be implemented in 2007. *Educational Sciences: Theory & Practice. 12*(3), 2230-2235.

Erwin, T. D., & Sebrell, K. W. (2003). Assessment of critical thinking: ETS's tasks in critical thinking. *The Journal of General Education*, *52*(1), 50-70.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*(3), 272-299.

Facione, P. A. (1990). *The California critical thinking skills test-college level technical report 1: Experimental validation and content validity*. Millbrae, CA: California Academic Press.

Falk, J. H. (2009). *Identity and the museum visitor experience*. Walnut Creek, CA: Left Coast Press.

Feinstein, N. W., & Kirchgasler, K. L. (2015). Sustainability in science education? How the Next Generation Science Standards approach sustainability, and why it matters. *Science Education, 99*(1), 121-144.

Field, A. (2005). *Discovering statistics with SPSS* (2nd Edition). London, UK: Sage.

Foote, C. J. (1998). Student-generated higher order questioning as a study strategy. *The Journal of Educational Research*, *92*(2), 107-113.

Ford, M. (2008). Disciplinary authority and accountability in scientific practice and learning. *Science Education*, *92*(3), 404-423.

Fornell, C. & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research,* 39-50.

Fortus, D., & Vedder-Weiss, D. (2014). Measuring students' continuing motivation for science learning. *Journal of Research in Science Teaching*, *51*(4), 497-522.

Fredricks, J. A., & Eccles, J. S. (2002). Children's competence and value beliefs from childhood through adolescence: growth trajectories in two male-sex-typed domains. *Developmental Psychology, 38*(4), 519-533.

Gardner, G. T., & Stern, P. C. (2002). *Environmental problems and human behavior* (2nd ed.). Boston, MA: Pearson.

Gee, J. P. (2000). Identity as an analytic lens for research in education. *Review of Research in Education*, *25,* 99-125.

Gee, J. P. (2007). *What video games have to teach us about learning and literacy*, 2nd Edition. New York: Palgrave Macmillan.

Gehlbach, H., & Brinkworth, M. E. (2011). Measure twice, cut down error: A process for enhancing the validity of survey scales. *Review of General Psychology, 15*(4), 380-387.

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

Gignac, G. E. (2009). Partial confirmatory factor analysis: Described and illustrated on the NEO–PI–R. *Journal of Personality Assessment*, *91*(1), 40-47.

Gilbert, J. K., Bulte, A. M., & Pilot, A. (2011). Concept development and transfer in context-based science education. *International Journal of Science Education, 33*(6), 817-837.

Gilligan, C. (1982). *In a different voice.* Harvard University Press.

Glaser, B. S., & Strauss, A. A. (1967). *The discovery of grounded theory.* New York, NY: Aldine De Gruyter.

Glorfeld, L.W. (1995). An improvement on Horn's parallel analysis methodology for selecting the correct number of factors to retain. *Educational and Psychological Measurement, 55*, 377-393.

Glynn, S. M., Brickman, P., Armstrong, N., & Taasoobshirazi, G. (2011). Science motivation questionnaire II: Validation with science majors and nonscience majors. *Journal of Research in Science Teaching, 48*(10), 1159-1176.

Glynn, S. M., & Koballa, T. R. (2006). Motivation to learn in college science. In J. J. Mintzes & W. H. Leonard (Eds.), *Handbook of college science teaching* (pp. 25-32). Arlington, VA: National Science Teachers Association Press.

Golafshani, N. (2003). Understanding reliability and validity in qualitative research. *The Qualitative Report*, *8*(4), 597-606.

Goldman, D., Yavetz, B., & Peer, S. (2006). Environmental literacy in teacher training in Israel: Environmental behavior of new students. *Journal of Environmental Education, 38*(1), 3-22. http://dx.doi.org/10.3200/JOEE.38.1.3-22

Gorsuch, R. L. (2003). Factor analysis. In J. A. Schinka & F. Velicer (Eds.), *Handbook of psychology: Vol. 2. Research Methods in Psychology* (pp. 143-164). Hoboken, NJ: John Wiley.

Gothwal, V. K., Wright, T. A., Lamoureux, E. L., & Pesudovs, K. (2009). Rasch analysis of the quality of life and vision function questionnaire. *Optometry & Vision Science, 86*(7), 836-844.

Gough, A. (1998). Researching environmental education in teacher education: Initiating and sustaining student interest. *International Research in Geographical and Environmental Education, 7*(3), 260-264.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review, 101,* 371–395. doi:10.1037/0033-295x.101.3.371

Graesser, A. C., & Person, N. K. (1994). Question asking during tutoring. *American Educational Research Journal*, *31*(1), 104-137.

Graesser, A. C., Lu, S., Olde, B. A., Cooper-Pye, E., & Whitten, S. (2005). Question asking and eye tracking during cognitive disequilibrium: Comprehending illustrated texts on devices when the devices break down. *Memory & Cognition, 33*(7), 1235-1247.

Graham, S. (2000). *Academic achievement and ethnic minority students: Constraints and opportunities.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Graham, S., & Weiner, B. (1996). Theories and principles of motivation. *Handbook of Educational Psychology, 4*, 63-84.

Gresalfi, M. S. (2004). Taking up opportunities to learn: Examining the construction of participatory mathematical identities in middle school classrooms. *Unpublished Dissertation*, Stanford, Palo Alto.

Gresalfi, M. S. (2009). Taking up opportunities to learn: Constructing dispositions in mathematics classrooms. *The Journal of the Learning Sciences*, *18*(3), 327-369.

Grigg, W. S., Lauko, M. A., & Brockway, D. M. (2006). *The nation's report card: Science 2005* (NCES 2006–466). Washington, DC: U.S. Government Printing Office.

Grinnell, F. (2013). Research integrity and everyday practice of science. *Science and Engineering Ethics*, *19*(3), 685-701.

Gross, M. (2001). Imposed information seeking in public libraries and school library media centers: a common behavior. *Information Research*, *6*(2), 6-2.

Groves, R. M., Fowler Jr, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey methodology* (Vol. 561). Hoboken, NJ: John Wiley & Sons.

Guzdial, M., & Turns, J. (2000). Effective discussion through a computer-mediated anchored forum. *The Journal of the Learning Sciences*, *9*(4), 437-469.

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2010). *Multivariate Data Analysis* (7th Eds.). New York, NY: Pearson.

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

Hambleton, R., Swaminathan, H., Rogers, H. (1991). *Fundamentals of Item Response Theory*. Newbury Park, California: Sage Publications, Inc.

Harris, J. (2005). Our agenda for technology integration: It's time to choose. *Contemporary Issues in Technology and Teacher Education*, *5*(2).

Harter, S. (1988). Developmental processes in the construction of the self. In T. D. Yawkey & T. E. Johnson (Eds.), *Integrative processes and socialization: Early to middle childhood* (pp. 45-78). Hillsdale, NJ: Erlbaum.

Hatt-Echeverria, B. (2005). *What does it mean to be smart? The figured world of smartness and student achievement.* Paper presented at the annual meeting of the American Educational Research Association, Montreal, CA.

Heilbronner, N. N. (2011). Stepping onto the STEM pathway: Factors affecting talented students' declaration of STEM majors in college. *Journal for the Education of the Gifted*, *34*, 876–899.

Henson, R. K., & Roberts, J. K. (2006). Use of exploratory factor analysis in published research common errors and some comment on improved practice. *Educational and Psychological Measurement*, *66*(3), 393-416.

Hibbard, J. H., Mahoney, E. R., Stockard, J., & Tusler, M. (2005). Development and testing of a short form of the patient activation measure. *Health Services Research*, *40*(6p1), 1918-1930.

Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist, 41*(2), 111-127.

Hidi, S., & Baird, W. (1986). Interestingness - A neglected variable in discourse processing. *Cognitive Science, 10*(2), 179-194.

Hill, C., Corbett, C., & St Rose, A. (2010). *Why so few? Women in science, technology, engineering, and mathematics.* American Association of University Women. 1111 Sixteenth Street NW, Washington, DC 20036.

Hines, S. M. (2003). *Multicultural science education: Theory, practice, and promise*. New York: Peter Lang.

Hmelo, C. E., & Lin, X. (2000). Becoming self-directed learners: Strategy development in problem-based learning. In D. Evensen & C. E. Hmelo (Eds.), *Problem-based learning: A research perspective on learning interactions* (pp. 227-250). Mahwah, NJ: Erlbaum.

Hmelo-Silver, C. E., & Barrows, H. S. (2006). Goals and strategies of a problem-based learning facilitator. *The Interdisciplinary Journal of Problem-Based Learning*, *1*(1), 4.

Hmelo-Silver, C. E., & Barrows, H. S. (2008). Facilitating collaborative knowledge building. *Cognition and Instruction*, *26*(1), 48-94.

Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist*, *42*(2), 99-107.

Hofstein, A., & Lunetta, V. N. (2004). The laboratory in science education: Foundations for the 21st century, *Science Education, 88,* 28-54.

Horn, I. S. (2008). Turnaround students in high school mathematics: Constructing identities of competence through mathematical worlds. *Mathematical Thinking and Learning*, *10*(3), 201-239.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30,* 179-185.

Hortensius, L. (2012). *Project for introduction to multivariate statistics: Measurement invariance*. Retrieved from http://www.tc.umn.edu/~horte005/docs/MeasurementInvariance.pdf

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1-55.

Hulleman, C. S., Durik, A. M., Schweigert, S., & Harackiewicz, J. M. (2008). Task values, achievement goals, and interest: An integrative analysis. *Journal of Educational Psychology, 100*, 398-416.

Hungerford, H. R., Litherland, R. A., Volk, T. L., Ramsey, J. M., & Peyton, R. B. (2003). *Investigating and evaluating environmental issues and actions: Skill development program*. Champagin, IL: Stipes Publishing, LLC.

Ichikawa, M., & Konishi, S. (1995). Application of the bootstrap methods in factor analysis. *Psychometrika, 60*, 77–93.

Ismail, H. N., & Alexander, J. M. (2005). Learning within scripted and nonscripted peer-tutoring sessions: The Malaysian context. *The Journal of Educational Research, 99*(2), 67-77.

Jackson, D. L., Gillaspy Jr, J. A., & Purc-Stephenson, R. (2009). Reporting practices in confirmatory factor analysis: An overview and some recommendations. *Psychological Methods, 14*(1), 6-23.

Jacobs, J. E., Lanza, S., Osgood, D. W., Eccles, J. S., & Wigfield, A. (2002). Changes in children's self-competence and values: Gender and domain differences across grades one through twelve. *Child Development, 73*(2), 509-527.

Jagger, S. (2013). Affective learning and the classroom debate. *Innovations in Education and Teaching International, 50*(1), 38-50.

James, W. (1890). *Principles of Psychology*. New York: Holt.

Johnson, B., & Manoli, C. C. (2011). The 2-MEV scale in the United States: A measure of children's environmental attitudes based on the theory of ecological attitude. *The Journal of Environmental Education, 42*(2), 84-97.

Johnson, L. F., Levine, A., Smith, R. S., & Haywood, K. (2010). Key Emerging Technologies for Elementary and Secondary Education. *Tech Directions*, *70*(3), 33.

Joppe, M. (2000). *The research process.* Retrieved from http://www.htm.uoguelph.ca/pagefiles/MJResearch/ResearchProcess/home.html.

Jöreskog, K. G. (1967). A general approach to confirmatory maximum likelihood factor analysis. *ETS Research Bulletin Series*, 183-202.

Jovanovic, J., & King, S. S. (1998). Boys and Girls in the Performance-Based Science classroom: Who's doing the performing? *American Educational Research Journal, 35*(3), 477-496.

Jowett, T., Harraway, J., Lovelock, B., Skeaff, S., Slooten, L., Strack, M., & Shephard, K. (2014). Multinomial-regression modeling of the environmental attitudes of higher education students based on the revised new ecological paradigm scale. *The Journal of Environmental Education, 45*(1), 1-15.

Kadlec, A., Friedman, W., & Ott, A. (2007). *Important, but not for me: Kansas and Missouri students and parents talk about math, science, and technology education.* New York: Public Agenda. Retrieved from http://files.eric.ed.gov/fulltext/ED498649.pdf.

Kaiser, H.F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement, 20,* 141-151.

Källkvist, M., Gomez, S., Andersson, H., & Lush, D. (2009). Personalised virtual learning spaces to support undergraduates in producing research reports: Two case studies. *The Internet and Higher Education*, *12*(1), 35-44.

Kane, J. M. (2012). Young African American children constructing academic and disciplinary identities in an urban science classroom. *Science Education*, *96*(3), 457-487.

Kane, J. M. (2016). Young African American boys narrating identities in science. *Journal of Research in Science Teaching*, *53*(1), 95-118.

Kennedy, C., Wilson, M. R., Draney, K., Tutunciyan, S., & Vorp, R. (2011). *ConstructMap Downloads.* Retrieved from http://bearcenter.berkeley.edu/ConstructMap/download.php.

Ketelhut, D. J., & Nelson, B. C. (2010). Designing for real-world scientific inquiry in virtual environments. *Educational Research*, *52,* 151–167.

King, A. (1989). Effects of self-questioning training on college students' comprehension of lectures. *Contemporary Educational Psychology*, *14*(4), 366-381.

King, A. (1990). Enhancing peer interaction and learning in the classroom through reciprocal questioning. *American Educational Research Journal*, *27*(4), 664-687.

King, A. (1991). Effects of training in strategic questioning on children's problem-solving performance. *Journal of Educational Psychology*, *83*(3), 307.

King, A. (1992). Comparison of self-questioning, summarizing, and notetaking-review as strategies for learning from lectures. *American Educational Research Journal*, *29*(2), 303-323.

King, A. (1992a). Facilitating elaborative learning through guided student-generated questioning. *Educational Psychologist*, *27*(1), 111-126.

King, A. (1994). Autonomy and question asking: The role of personal control in guided student-generated questioning. *Learning and Individual Differences*, *6*(2), 163-185.

King, A. (1995). Designing the instructional process to enhance critical thinking across the curriculum. Inquiring minds really do want to know: Using questioning to teach critical thinking. *Teaching of Psychology*, *22*(1), 13-17.

King, A., & Rosenshine, B. (1993). Effects of guided cooperative questioning on children's knowledge construction. *Journal of Experimental Education*, *61*, 127–148.

Kline, P. (2014). *An easy guide to factor analysis.* New York, NY: Routledge.

Kline, R. B. (2005). *Principles and Practice of Structural Equation Modeling*, (2nd ed.). New York: Guilford.

Koepp, M. J., Gunn, R. N., Lawrence, A. D., Cunningham, V. J., Dagher, A., Jones, T., & Grasby, P. M. (1998). Evidence for striatal dopamine release during a video game. *Nature*, *393*(6682), 266-268.

Kollar, I., Fischer, F., & Hesse, F. W. (2006). Collaboration scripts–a conceptual analysis. *Educational Psychology Review*, *18*(2), 159-185.

Kramarski, B., & Dudai, V. (2009). Group-metacognitive support for online inquiry in mathematics with differential self-questioning. *Journal of Educational Computing Research*, *40*, 377–404.

Krapp, A., Hidi, S., & Renninger, K. A. (1992). Interest, learning and development. In K. A. Renninger, S. Hidi, & A. Krapp (Eds.), *The role of interest in learning and development* (pp. 3-25). Hillsdale, NJ: Erlbaum.

Krosnick, J. A., & Petty, R. E. (1995). Attitude strength: An overview. In R. E. Petty & J. A. Krosnick (Eds.), *Attitude strength: Antecedents and consequences* (Vol. 4, pp. 1–24). Hillsdale, NJ: Erlbaum.

Kuhn, D. (2007). Is direct instruction an answer to the right question? *Educational Psychologist*, *42*(2), 109-113.

Kuhn, L., & Reiser, B. (2006). *Structuring activities to foster argumentative discourse*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Kulas, J. T., Stachowski, A. A., & Haynes, B. A. (2008). Middle response functioning in Likert-responses to personality items. *Journal of Business and Psychology*, *22*(3), 251-259.

Kwon, K., Hong, R. Y., & Laffey, J. M. (2013). The educational impact of metacognitive group coordination in computer-supported collaborative learning. *Computers in Human Behavior*, *29*(4), 1271-1281.

La Guardia, J. G. (2009). Developing who I am: A self-determination theory approach to the establishment of healthy identities. *Educational Psychologist*, *44*(2), 90-104.

Lacasa, P., Martínez, R., & Del Castillo, H. (2011). *Adolescent thinking and online writing after the use of commercial games in the classroom.* Retrieved from http://www.digra.org/wp-content/uploads/digital-library/11312.53222.pdf

Lang, S. (2007). *Water, air and soil pollution causes 40 percent of deaths worldwide, Cornell research survey finds.* Retrieved from http://www.news.cornell.edu/stories/aug07/morediseases.sl.html

Lave, J. (1992, April). *Learning as participation in communities of practice*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. New York: Cambridge University Press.

Lee, J. (2004). Evaluating the effectiveness of instructional resource allocation and use: IRT and HLM analysis of NAEP teacher survey and student assessment data. *Studies in Educational Evaluation*, *30*(2), 175-199.

Lent, R. W., Lopez, F. G., Brown, S. D., & Gore Jr, P. A. (1996). Latent structure of the sources of mathematics self-efficacy. *Journal of Vocational Behavior*, *49*(3), 292-308.

Levy, F., & Murnane, R. J. (2012). *The new division of labor: How computers are creating the next job market*. Princeton, NJ: Princeton University Press.

Lewin, K. (1938). *The conceptual representation and the measurement of psychological forces*. Durham, NC: Duke University Press.

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

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, *140*, 5-55.

Linacre, J. M. (2012). *Winsteps ministep. Rasch-model computer program.* Retrieved from http://www.winsteps.com/winman/

Linacre, J. M. (2014). *Facets Rasch measurement computer program* (Version 3.71.4)[Computer software]. Chicago, IL: Winsteps.com.

Lipman, P. (2004). *High stakes education: Inequality, globalization, and urban school reform*. New York, NY: Routledge.

Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling*, *9*(2), 151-173.

Little, T. D., Rhemtulla, M., Gibson, K., & Schoemann, A. M. (2013). Why the items versus parcels controversy needn't be one. *Psychological Methods*, *18*(3), 285.

Liu, L., & Hmelo-Silver, C. E. (2009). Promoting complex systems learning through the use of conceptual representations in hypermedia. *Journal of Research in Science Teaching*, *46*(9), 1023-1040.

Lloyd, B., & Duveen, G. (1992). *Gender identities and education: The impact of starting school*. New York: St. Martin's Press.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. New York, NY: Routledge.

Lord, R. T. (1999). A comparison between traditional and constructivist teaching in environmental science. *The Journal of Environmental Education*, *30*(3), 22-28.

MacCallum, R. C., & Austin, J. T. (2000). Applications of structural equation modeling in psychological research. *Annual Review of Psychology*, *51*(1), 201-226.

MacKnight, C. B. (2000). Teaching critical thinking through online discussions. *Educause Quarterly*, *23*(4), 38-41

Mallinson, T. (2007). Why measurement matters for measuring patient vision outcomes. *Optometry & Vision Science*, *84*(8), 675-682.

Mallinson, T., Stelmack, J., & Velozo, C. (2004). A comparison of the separation ratio and coefficient α in the creation of minimum item sets. *Medical Care*, *42*(1), 1-17.

Marsh, H. W. (1996). Positive and negative global self-esteem: A substantively meaningful distinction or artifactors? *Journal of Personality and Social Psychology*, *70*(4), 810.

Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin*, *103*(3), 391.

Matsui, T., Matsui, K., & Ohnishi, R. (1990). Mechanisms underlying math self-efficacy learning of college students. *Journal of Vocational Behavior*, *37*(2), 225-238.

Matsunaga, M. (2010). How to factor-analyze your data right: Do's, don'ts, and how-to's. *International Journal of Psychological Research*, *3*(1), 97-110.

Mayo, M. J. (2007). Games for science and engineering education. *Communications of the ACM*, *50*(7), 30-35.

McClelland, D. C. (1961). *The achieving society.* Princeton, NJ: Van Nostrand.

McCrum-Gardner, E. (2008). Which is the correct statistical test to use? *British Journal of Oral and Maxillofacial Surgery*, *46*(1), 38-41.

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

MetaMetrics, Inc. (2012). *The lexile framework for reading.* Retrieved from www.lexile.com.

Milfont, T. L., & Fischer, R. (2010). Testing measurement invariance across groups: Applications in cross-cultural research. *International Journal of Psychological Research*, *3*(1), 111-130.

Miri, B., David, B. C., & Uri, Z. (2007). Purposely teaching for the promotion of higher-order thinking skills: A case of critical thinking. *Research in Science Education*, *37*(4), 353-369.

Mitchell, M. (1993). Situational interest: Its multifaceted structure in the secondary school mathematics classroom. *Journal of Educational Psychology*, *85*(3), 424.

Moje, E. B. (2004). Powerful spaces: Tracing the out-of-school literacy spaces of Latino/a youth. In K. Leander & M. Sheehy (Eds.), *Spatializing literacy research and practice* (pp. 15-38). New York, NY: Peter Lang.

Moje, E. B., Tucker-Raymond, E., Varelas, M., & Pappas, C. C. (2007). FORUM: Giving oneself over to science: Exploring the roles of subjectivities and identities in learning science. *Cultural Studies of Science Education*, *1*(3), 593-601.

Moskal, B. M., Leydens, J. A., & Pavelich, M. J. (2002). Validity, reliability and the assessment of engineering education. *Journal of Engineering Education*, *91*(3), 351-354.

Mubeen, S., & Reid, N. (2014). The measurement of motivation with science students. *European Journal of Educational Research*, *3*(3), 129-144.

Muuss, R. (1996). *Theories of adolescence (6th ed.)*. New York, NY: McGraw Hill.

Naaz, F., Chariker, J. H., & Pani, J. R. (2014). Computer-based learning: Graphical integration of whole and sectional neuroanatomy improves long-term retention. *Cognition and Instruction*, *32*(1), 44-64.

National Research Council. (1996). *National Science Education Standards.* Washington, DC: National Academy Press.

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

National Science Foundation. (2001). *Characteristics of doctoral scientists and engineers in the United States. Detailed statistical tables.* Retrieved from https://evergreen.lib.in.us/eg/opac/record/17367200

Needham, D. R., & Begg, I. M. (1991). Problem-oriented training promotes spontaneous analogical transfer: Memory-oriented training promotes memory for training. *Memory & Cognition*, *19*(6), 543-557.

Nesselroade, J. R. (1994). Exploratory factor analysis with latent variables and the study of processes of development and change. In A. von Eye & C. C. Clogg (Eds.), *Analysis of latent variables in developmental research* (pp. 131-154). Newbury Park, CA: Sage.

Nevitt, J., & Hancock, G. R. (2001). Performance of bootstrapping approaches to model test statistics and parameter standard error estimation in structural equation modeling. *Structural Equation Modeling*, *8*(3), 353-377.

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

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instrumentation, and Computers, 32*(3), 396-402.

Osbaldiston, R. (2004). Meta-analysis of the responsible environmental behavior literature (Doctoral dissertation, University of Missouri—Columbia, 2004). *Dissertation Abstracts International*, *65*(8), 4340B.

Osborne, J. (2014). Teaching scientific practices: Meeting the challenge of change. *Journal of Science Teacher Education, 25(2), 177–196.*

Osborne, J., Erduran, S., & Simon, S. (2004). Enhancing the quality of argumentation in school science. *Journal of Research in Science Teaching, 41*(10), 994–1020.

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.

Otfinowski, R., & Silva-Opps, M. (2015). Writing toward a scientific identity: Shifting from prescriptive to reflective writing in undergraduate biology. *Journal of College Science Teaching*, *45*(2), 19.

Oyserman, D., Bybee, D., & Terry, K. (2006). Possible selves and academic outcomes: How and when possible selves impel action. *Journal of Personality and Social Psychology*, *91*(1), 188.

Palincsar, A. S. (1998). Social constructivist perspectives on teaching and learning. *Annual Review of Psychology*, *45*, 345–375.

Pea, R.D. (1993). Learning scientific concepts through material and social activities: Conversational analysis meets conceptual change. *Educational Psychology*, *28*, 265–277.

Pearson, P. D., Moje, E., & Greenleaf, C. (2010). Literacy and science: Each in the service of the other. *Science*, *328*(5977), 459–463.

Peters, V. L., & Hewitt, J. (2010). An investigation of student practices in asynchronous computer conferencing courses. *Computers & Education*, *54*(4), 951-961.

Peterson, R. A., & Kim, Y. (2013). On the relationship between coefficient alpha and composite reliability. *Journal of Applied Psychology, 98*(1), 194.

Pett, M. A., Lackey, N. R., & Sullivan, J. J. (2003). *Making sense of factor analysis: The use of factor analysis for instrument development in health care research*. Thousand Oaks, CA: Sage.

Pink, D. H. (2006). *A whole new mind: Why right-brainers will rule the future*. New York, NY: Penguin.

Pintrich, P. R. (2003). A motivational science perspective on the role of student motivation in learning and teaching contexts. *Journal of Educational Psychology, 95*(4), 667-686.

Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, *82*(1), 33-40.

Pintrich, P. R., & Schunk, D. H. (2002). *Motivation in education: Theory, research, and applications*. Englewood Cliffs, NJ: Prentice Hall.

PISA. (2012). *Results in Focus: What 15-year-olds know and what they can do with what they know.* Retrieved from http://www.oecd.org/pisa/keyfindings/pisa-2012-results-overview.pdf

Plante, I., O'Keefe, P. A., & Théorêt, M. (2013). The relation between achievement goal and expectancy-value theories in predicting achievement-related outcomes: A test of four theoretical conceptions. *Motivation and Emotion*, *37*(1), 65-78.

Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*, *104*(1), 1-15.

Rafal, C. T. (1996). From co-construction to takeovers: Science talk in a group of four girls. *Journal of Learning Sciences, 5*, 279–293.

Rasch, G. (1960). *Probabilistic models for some intelligence and achievement tests.* Copenhagen: Danish Institute for Educational Research.

Rasch, G. (1980). *Probabilistic Models for Intelligence and Attainment Tests.* Chicago, IL: University of Chicago Press.

Ratcliffe, M. (1997). Pupil decision-making about socio-scientific issues within the science curriculum. *International Journal of Science Education*, *19*(2), 167–182.

Ray, J. V., Frick, P. J., Thornton, L. C., Steinberg, L., & Cauffman, E. (2016). Positive and negative item wording and its influence on the assessment of callous-unemotional traits. *Psychological Assessment*, *28*(4), 394.

Rea, L. M., & Parker, R. A. (2014). *Designing and conducting survey research: A comprehensive guide*. San Francisco, CA: John Wiley & Sons.

Reise, S. P., Waller, N. G., & Comrey, A. L. (2000). Factor analysis and scale revision. *Psychological Assessment*, *12*(3), 287.

Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. *Psychological Bulletin*, *114*(3), 552.

Renninger, A., Hidi, S., & Krapp, A. (2014). *The role of interest in learning and development.* New York, NY: Psychology Press.

Renninger, K. A. (2000). Individual interest and its implications for understanding intrinsic motivation. In C. Sansone & J. M. Harackiewicz (Eds.), *Intrinsic and extrinsic motivation: The search for optimal motivation and performance* (pp. 375–407). New York: Academic.

Renninger, K. A. (2009). Interest and identity development in instruction: An inductive model. *Educational Psychologist, 44*(2), 105-118.

Roscoe, R. D., & Chi, M. T. (2008). Tutor learning: The role of explaining and responding to questions. *Instructional Science*, *36*(4), 321-350.

Rosenshine, B., Meister, C., & Chapman, S. (1996). Teaching students to generate questions: A review of the intervention studies. *Review of Educational Research, 66*(2), 181-221.

Rubin, B. C. (2007). Learner identity amid figured worlds: Constructing (in) competence at an urban high school. *The Urban Review*, *39*(2), 217-249.

Russell, D. W. (2002). In search of underlying dimensions: The use (and abuse) of factor analysis in Personality and Social Psychology Bulletin. *Personality and Social Psychology Bulletin, 28*(12), 1629-1646.

Russell, D. W., Lucas, K. B., & McRobbie, C. J. (2004). Role of the microcomputer-based laboratory display in supporting the construction of new understandings in thermal physics. *Journal of Research in Science Teaching*, *41*(2), 165-185.

Rutherford, F. J. (1997, October). Sputnik and science education. In symposium, *Reflecting on Sputnik: Linking the past, present, and future of educational reform*. Washington, DC.

Ryan, F. L. (1971). *Exemplars for the new social studies: Instructing in the elementary school*. Englewood Cliffs, NJ: Prentice Hall.

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, *25*(1), 54-67.

Ryoo, K. (2015). Teaching science through the language of students in technology-enhanced instruction. *Journal of Science Education and Technology*, *24*(1), 29-42.

Saavedra, A. R., & Opfer, V. D. (2012). Learning 21st-century skills requires 21st-century teaching. *Phi Delta Kappan*, *94*(2), 8-13.

Sadler, T.D. (2009). Situated learning in science education: Socio-scientific issues as contexts for practice. *Studies in Science Education*, *45*, 1-42.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7*(2), 147-177.

Scheuer, O., McLaren, B. M., Weinberger, A., & Niebuhr, S. (2014). Promoting critical, elaborative discussions through a collaboration script and argument diagrams. *Instructional Science*, *42*(2), 127-157.

Schibeci, R. A. (1984). Attitudes to science: An update. *Studies in Science Education, 11*, 26-59.

Schibeci, R. A., & Riley, J. P. (1986). Influence of students' background and perceptions on science attitudes and achievement. *Journal of Research in Science Teaching, 23*(3), 177-187.

Schiefele, U. (1992). Topic interest and levels of text comprehension. In K. A. Renninger, S. Hidi, & A. Krapp (Eds.), *The role of interest in learning and development* (pp. 151–183). Hillsdale, NJ: Erlbaum.

Schiefele, U. (2009). Situational and individual interest. In K. R. Wentzel & A. Wigfield (Eds.), *Handbook of motivation at school* (pp. 197-222). New York, NY: Taylor Francis.

Schlecher, A. (Ed.). (2012). *Preparing teachers and developing school leaders for the 21st century: Lessons from around the world.* Paris, France: OECD Publishing.

Schraw, G., & Lehman, S. (2001). Situational interest: A review of the literature and directions for future research. *Educational Psychology Review, 13*(1), 23-52.

Schriesheim, C. A., Eisenbach, R. J., & Hill, K. D. (1991). The effect of negation and polar opposite item reversals on questionnaire reliability and validity: An experimental investigation. *Educational and Psychological Measurement*, *51*(1), 67-78.

Schulz, W., & Fraillon, J. (2011). The analysis of measurement equivalence in international studies using the Rasch model. *Educational Research and Evaluation, 17*(6), 447-464.

Schunk, D. H., Pintrich, P. R., & Meece, J. L. (2008). *Motivation in education: Theory, research, and applications.* New Jersey: Pearson.

Schwartz, D. L., Blair, K. P., Biswas, G., Leelawong, K., & Davis, J. (2007). Animations of thought: Interactivity in the teachable agents paradigm. In R. Lowe & W. Schnotz (Eds.), *Learning with animation: Research and implications for design* (pp. 114-140). UK: Cambridge University Press.

Schwartz, D. L., Bransford, J. D., & Sears, D. (2005). Efficiency and innovation in transfer. In J. P. Mestre (Ed.), *Transfer of learning from a modern multidisciplinary perspective* (pp. 1-51). Greenwich, CT: Information Age.

Schwartz, M. S., & Fischer, K. W. (2006). Useful metaphors for tackling problems in teaching and learning. *About Campus*, *11*(1), 2-9.

Sfard, A., & Prusak, A. (2005). Telling identities: In search of an analytic tool for investigating learning as a culturally shaped activity. *Educational Researcher*, *34*(4), 14-22.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Belmont, CA: Wadsworth Cengage Learning.

Shields, P. M., & Rangarajan, N. (2013). *A playbook for research methods: Integrating conceptual frameworks and project management.* Stillwater, OK: New Forums Press.

Silva, E. (2009). Measuring skills for 21st century learning. *The Phi Delta Kappan, 90*(9), 630-634.

Silvia, P. J. (2001). Interest and interests: The psychology of constructive capriciousness. *Review of General Psychology, 5*, 270–290.

Silvia, P. J. (2005). What is interesting? Exploring the appraisal structure of interest. *Emotion, 5*, 89-102.

Simon, S., & Amos, R. (2011). Decision making and use of evidence in a socio-scientific problem on air quality. In T. D. Sadler (Ed.), *Socio-scientific issues in the classroom* (Vol.39, pp. 167-192). Dordrecht, The Netherlands: Springer Netherlands.

Sjaastad, J. (2014). Enhancing measurement in science education research through Rasch analysis: Rationale and properties. *Nordic Studies in Science Education*, *10*(2), 212-230.

Slavin, R. E. (2008). Comprehensive school reform. In T. Good et al. (Eds.), *21st century education: A reference handbook.* Thousand Oaks, CA: Sage.

Sobel, D. (1996). *Beyond ecophobia: Reclaiming the heart in nature education.* Great Barrington, MA: The Orion Society and The Myrin Institute.

Spector, P. E., Van Katwyk, P. T., Brannick, M. T., & Chen, P. Y. (1997). When two factors don’t reflect two constructs: How item characteristics can produce artifactual factors. *Journal of Management*, *23*(5), 659-677.

Stark, R., & Gray, D. (1999). Gender preferences in learning science. *International Journal of Science Education*, *21*(6), 633-643.

Steele, C. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, *52*, 613–629.

Steiger, J. H., & Lind, J. C. (1980, May). *Statistically based tests for the number of common factors.* Paper presented at the annual meeting of the Psychometric Society, Iowa City, IA.

Stenner, A. J., Burdick, H., Sanford, E. E., & Burdick, D. S. (2005). How accurate are Lexile text measures? *Journal of Applied Measurement*, *7*(3), 307-322.

Stern, J. M., Powell, B. R., & Ardoin, M. N. (2011). Evaluating a constructivist and culturally responsive approach to environmental education for diverse audiences. *The Journal of Environmental Education, 42*(2), 109-122.

Stetsenko, A. (2008). From relational ontology to transformative activist stance on development 153 and learning: Expanding Vygotsky's (CHAT) project. *Cultural Studies of Science Education, 3*, 471-491.

Stevens, J. (1996). *Applied multivariate statistics for the social sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Stevens, T., Olivárez, A., & Hamman, D. (2006). The role of cognition, motivation, and emotion in explaining the mathematics achievement gap between Hispanic and White students. *Hispanic Journal of Behavioral Sciences, 28*(2), 161-186.

Stone, E. F. (1978). *Research methods in organizational behavior.* Santa Monica, CA: Goodyear Publishing Company.

Stryker, S. (1980). *Symbolic interactionism: A social structural version*. Menlo Park, CA: Benjamin-Cummings.

Swan, A. E., & O'Donnell, A. M. (2009). The contribution of a virtual biology laboratory to college students' learning. *Innovations in Education and Teaching International, 46*(4), 405-419.

Swanson, C. B. (2004). *Graduation rates: Real kids, real numbers.* Washington, DC: The Urban Institute.

Swarat, S., Ortony, A., & Revelle, W. (2012). Activity matters: Understanding student interest in school science. *Journal of Research in Science Teaching*, *49*(4), 515-537.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics.* Boston, MA: Allyn and Bacon.

Tabak, I., & Reiser, B. J. (2008). Software-realized inquiry support for cultivating a disciplinary stance. *Pragmatics & Cognition*, *16*(2), 307-355.

Tan, E., & Barton, A. C. (2008). Unpacking science for all through the lens of identities-in-practice: The stories of Amelia and Ginny. *Cultural Studies of Science Education*, *3*(1), 43-71. Also available at http://barton.wiki.educ.msu.edu/file/view/tan+and+ACB.pdf

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education, 2*, 53-55.

Tennant, A., & Conaghan, P. G. (2007). The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis Care & Research*, 57(8), 1358-1362.

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

Toledo, C. A. (2015). Dog bite reflections - Socratic questioning revisited. *International Journal of Teaching and Learning in Higher Education*, *27*(2), 275-279.

Tolman, E. C. (1932). *Purposive behavior in men and animals*. New York: Appleton-Century- Crofts.

Trochim, W. M. (2006). *The Research Methods Knowledge Base*, 2nd Edition. Retrieved from http://www.socialresearchmethods.net/kb/.

Tsagkanos, A. G. (2007). A bootstrap-based minimum bias maximum simulated likelihood estimator of mixed logit. *Economics Letters*, *96*(2), 282-286.

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, *38*, 1-10.

Tucker-Raymond, E., Varelas, M., Pappas, C. C., Korzh, A., & Wentland, A. (2007). “They probably aren’t named Rachel”: Young children’s scientist identities as emergent multimodal narratives. *Cultural Studies of Science Education*, *1*(3), 559-592.

Tyler-Wood, T., Knezek, G., & Christensen, R. (2010). Instruments for assessing interest in STEM content and careers. *Journal of Technology and Teacher Education*, *18*(2), 345-368.

Tytler, R. (2007). Re-imagining science education: Engaging students in science for Australia. *Lab-talk*, *51*(3), 6-9. Also available at http://dro.deakin.edu.au/eserv/DU:30013528/tytler-reimaginingscienceeducation-2007.pdf

Usher, E. L., & Pajares, F. (2009). Sources of self-efficacy in mathematics: A validation study. *Contemporary Educational Psychology*, *34*(1), 89-101.

Van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology*, *9*(4), 486-492.

Van Gelder, T. (2005). Teaching critical thinking: Some lessons from cognitive science. *College Teaching*, 41-46.

Varelas, M. (2012). *Identity construction and science education research: Learning, teaching, and being in multiple contexts* (Vol. 35). Springer Science & Business Media.

Venville, G., Adey, P., Larkin, S., Robertson, A., & Fulham, H. (2003). Fostering thinking through science in the early years of schooling. *International Journal of Science Education*, *25*(11), 1313-1331.

Verhelst, N. D., & Glas, C. A. (1993). A dynamic generalization of the Rasch model. *Psychometrika*, *58*(3), 395-415.

Vosniadou, S. (2013). Conceptual change in learning and instruction: The framework theory approach. *International Handbook of Research on Conceptual Change*, *2*, 11-30.

Wagner, J. F. (2006). Transfer in pieces. *Cognition and Instruction*, *24*(1), 1-71.

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

Wang, C. Y., Wu, H. K., Lee, S. W. Y., Hwang, F. K., Chang, H. Y., Wu, Y. T., & Tsai, C. C. (2014). A review of research on technology-assisted school science laboratories. *Educational Technology & Society*, *17*(2), 307-320.

Wang, W. C., Yao, G., Tsai, Y. J., Wang, J. D., & Hsieh, C. L. (2006). Validating, improving reliability, and estimating correlation of the four subscales in the WHOQOL-BREF using multidimensional Rasch analysis. *Quality of Life Research*, *15*(4), 607-620.

Warner, J. M. (2008). Supplemental instruction for non-science majors biology students: Meanings and influences on science identities for women. *Dissertation Abstracts International*, *69*(7), 2658A.

Watson, G. & Glaser, E. M. (1980). *Watson-Glaser critical thinking appraisal manual*: Forms A&B. San Antonio, TX: The Psychological Corporation.

Watts, M., Gould, G., & Alsop, S. (1997). Questions of understanding: Categorizing pupils' questions in science. *School Science Review*, *79*(286), 57–63.

Weaver-Hightower, M. (2003). The "boy turn" in research on gender and education. Review of *Educational Research*, *73*(4), 471-498.

Webb, N. M. (1989). Peer interaction, problem solving, and cognition: Multidisciplinary perspectives. *International Journal of Educational Research*, *13*, 1–119.

Wecker, C., Stegmann, K., Bernstein, F., Huber, M. J., Kalus, G., Kollar, I., & Fischer, F. (2010). S-COL: A Copernican turn for the development of flexibly reusable collaboration scripts. *International Journal of Computer-Supported Collaborative Learning*, *5*(3), 321-343.

Weinburgh, M. (1995). Gender differences in student attitudes toward science: A meta-analysis of the literature from 1970 to 1991. *Journal of Research in Science Teaching*, *32*(4), 387-398.

Weiner, B. (1992). *Human motivation: Metaphors, theories, and research*. Newbury Park, CA: Sage Publications.

Weller, J. A., Dieckmann, N. F., Tusler, M., Mertz, C. K., Burns, W. J., & Peters, E. (2013). Development and testing of an abbreviated numeracy scale: A Rasch analysis approach. *Journal of Behavioral Decision Making*, *26*(2), 198-212.

Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity*. Cambridge, UK: Cambridge University Press.

Wigfield, A., & Eccles, J. S. (1992). The development of achievement task values: A theoretical analysis. *Developmental Review*, *12*(3), 265-310.

Wigfield, A., Tonks, S., & Eccles, J. S. (2004). Expectancy value theory in cross-cultural perspective. In D. McInerney & S. van Etten (Eds.), *Research on sociocultural influences on motivation and learning* (pp. 165-198). Greenwich, CT: Information Age Publishers.

Wigfield, A., Tonks. S., & Klauda, S. L. (2009). Expectancy-value theory. In K. R.Wentzel & A. Wigfield (Eds.), *Handbook of motivation at school* (pp. 55-75). New York, NY: Routledge.

Williams, L. J., & O'Boyle, E. H. (2008). Measurement models for linking latent variables and indicators: A review of human resource management research using parcels. *Human Resource Management Review*, *18*(4), 233-242.

Williams, R. L., Oliver, R., & Stockdale, S. L. (2004). Psychological versus generic critical thinking as predictors and outcome measures in a large undergraduate human development course. *The Journal of General Education*, *53*(1), 37-58.

Wilson, M. (2005). *Constructing measures: An item response modeling approach.* Mahwah, NJ: Lawrence Erlbaum Associates.

Wimer, J. W., Ridenour, C. S., Thomas, K., & Place, A. W. (2001). Higher order teacher questioning of boys and girls in elementary mathematics classrooms. *The Journal of Educational Research*, *95*(2), 84-92.

Wright, B. D. (1992). Raw scores are not linear measures: Rasch vs. Classical Test Theory CTT Comparison. *Rasch Measurement Transactions*, *6*(1), 208. Retrieved from http://www.rasch.org/rmt/rmt61n.htm

Wright, B. D. (1997). A history of social science measurement. *Educational Measurement: Issues and Practice*, *16*(4), 33-45.

Wright, B. D., & Masters, G. N. (1982). *Rating Scale Analysis. Rasch Measurement.* Chicago, IL: MESA Press.

Wright, B. D., & Stone, M. H. (1979). *Best Test Design. Rasch Measurement*: Chicago, IL: Measurement, Evaluation, Statistics, and Assessment Press.

Wright, B. D., Linacre, J. M., Gustafson, J. E., & Martin-Lof, P. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, *8*(3), 370.

Wu, M. L., Adams, R. J., Wilson, M. R., & Haldane, S. (2007). *ACER ConQuest Version 2.0*. Manual Retrieved from http://www.rasch.org/rmt/rmt133o.htm.

Yarden, A., Brill, G., & Falk, H. (2001). Primary literature as a basis for a high-school biology curriculum. *Journal of Biological Education*, *35*, 190–195.

Zain, A. N. M., Samsudin, M. A., & Jusoh, A. (2010). Using the Rasch Model to Measure Students' Attitudes toward Science in" Low Performing" Secondary Schools in Malaysia. *International Education Studies*, *3*(2), 56-63.

Zecha, S. (2010). Environmental knowledge, attitudes and actions of Bavarian (southern Germany) and Asturian (northern Spain) adolescents. *International Research in Geographical and Environmental Education*, *19*(3), 227-240.

Zohar, A., & Dori, Y. J. (2003). Higher order thinking skills and low-achieving students: Are they mutually exclusive? *The Journal of the Learning Sciences*, *12*(2), 145-181.

Zoller, U. (1994). The examination where the student asks the questions. *School Science and Mathematics*, *94*, 347–349.

Zwick, W. R. & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin, 99*, 432-442.