DISSERTATION

DETERMINING THE ALIGNMENT BETWEEN WHAT TEACHERS ARE EXPECTED TO TEACH, WHAT THEY KNOW, AND HOW THEY ASSESS SCIENTIFIC LITERACY

Submitted by

Lisa Noel Pitot

School of Education

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Spring 2014

Doctoral Committee:

Advisor: Meena Balgopal

Donna Cooner
Karen Rambo-Hernandez
Pamela Coke

ABSTRACT

DETERMINING THE ALIGNMENT BETWEEN WHAT TEACHERS ARE EXPECTED TO TEACH, WHAT THEY KNOW, AND HOW THEY ASSESS SCIENTIFIC LITERACY

Science education reform efforts have highlighted the need for a scientifically literate citizen, capable of using their scientific knowledge and skills for reasoning, argumentation, and decision-making. Yet little is known about secondary science teachers' understandings of these reform efforts, specifically their knowledge, skills, and abilities related to scientific literacy. In addition to reform efforts, education policies have been enacted by states that rate a teachers' effectiveness in part on their students' performance on high stakes assessments. This study used multiple methods to examine a) teacher perceptions of scientific literacy, b) their scientific literacy, and c) their abilities to develop common science assessments that reflect scientific literacy skills defined by science education reformers. Using constant comparative analysis, open survey responses from secondary science teachers (n =48) from one district revealed that their perceptions of scientific literacy are not in alignment with those of science education reforms. Secondary science teachers (n =28) demonstrated as a group that they were scientifically literate. The assessments (n= 13) secondary science teachers developed did not align with the scientific literacy sub-constructs as defined by science education reformers. These finding inform the types of professional development that would benefit secondary science teachers.

ACKNOWLEDGEMENTS

Sincere gratitude is accredited to my advisor and mentor—Dr. Meena Balgopal, without you I would never have been able to finish this dissertation. You knew when to be a gentle guide and when to crack the whip, and most important your intelligence and keen understanding of science education helped me put my findings into perspective. In addition, my sincere thanks to our Science Education Research Group members—Aramati, Leila, Katie, and Neely—Go SERG!

I am also grateful to my committee members, all of whom provided their expertise and feedback to shape my dissertation into this final product.

I acknowledge my colleagues from—the School of Education and the Department of Natural Resources at Colorado State University, the Poudre Learning Center, and my fellow educators in Poudre School District—whether it be in a class that we took together, a professional development experience we created for teachers, or the day in and day out job of teaching kids—I have learned and continue to learn so much from all of you. A special thanks goes out to my colleagues who took the time to complete my survey.

DEDICATION

I dedicate my dissertation to my family. My sons James and Lars—Thank-you for your patience and the sacrifices you made while I have been on this journey, I hope that my example of perseverance will be a guide for you as you both make your marks on this world. To my parents—Julie and Henry-- Thank-you for your support and believing in me not just on this adventure, but on the many I've taken through the years that have gotten me to this point. To my sisters and brother—Anita, Jeanne, Cathy, Henry, Michelle, and Patrice—you all motivate me to greatness and I thank-you for believing in me. And lastly, I dedicate my writing to our sister Bertha Elizabeth—the angel among us—who lived her life with courage and faith, thank-you for your inspiration.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

DEFINITION OF TERMS / ACRONYMS

Alignment: The degree to which the components of an educational system work together to achieve the desired goals of stakeholders (Case and Zucker, 2008).

Claim: "A claim is a statement that expresses the answer or conclusion to a question" (McNeill & Krajcik, 2012, p. 22). "An assertion that is based on evidence or knowledge" (The College Board, 2009, p. 203).

Content Analysis: The systematic, replicable technique for compressing many words of text into fewer content categories based on explicit rules of coding (Krippendorff, 1980; Nuendorf, 2002).

Evidence: "Evidence is scientific data that supports the claim. Data are information such as observations and measurements that come from natural settings" (McNeill & Krajcik, 2012, p. 23). Evidence is data "that have been represented, analyzed, and interpreted in the context of a specific scientific question" (The College Board, 2009, p. 204).

NSES: National Science Education Standards

NGSS: Next Generation Science Standards

NOS: Nature of Science

NRC: National Research Council

QR: Quantitative Reasoning also known as Numeracy or Quantitative Literacy: "A *habit of mind*, competency, and comfort in working with numerical data. Individuals with strong QR skills possess the ability to reason and solve quantitative problems from a wide array of authentic contexts and everyday life situations. They understand and can create sophisticated arguments supported by quantitative evidence and they can clearly communicate those arguments in a variety of formats (using words, tables, graphs, mathematical equations, etc., as

appropriate)" (Association of American Colleges and Universities, 2012, par.2).

Reasoning: "The reasoning explains why the evidence supports the claim, providing a logical explanation between the evidence and claim. It typically requires the discussion of appropriate scientific principles to explain that link" (McNeill & Krajcik, 2012, p. 24)

Scientific Explanation: "Includes an assertion (claim) about natural systems or designed objects, or phenomena; the evidence, which can consist of empirical evidence or observations, related to the claim; and reasoning (argumentation that links the claim with the evidence" (The College Board, 2009, p.206).

SL - Scientific Literacy: "Scientific literacy is the knowledge and understanding of scientific concepts and processes required for personal decision making, participation in civic and cultural affairs, and economic productivity… A literate citizen should be able to evaluate the quality of scientific information on the basis of its source and the methods used to generate it. Scientific literacy also implies the capacity to pose and evaluate arguments based on evidence and to apply conclusions from such arguments appropriately" (National Research Council. 1996).

SSI – Socio-scientific issue: Science-based societal issue or problem that is "controversial in nature, discussed in public outlets, and frequently subject to political and ethical influences" (Sadler & Zeidler, 2005, p. 113).

TOSLS - Test of science literacy skills: Valid instrument (28-item multiple choice) developed to measure the scientific literacy of undergraduate biology majors. (Gormally, Brickman, & Lutz, 2012).

CHAPTER 1: INTRODUCTION

Currently, humans are facing a multitude of scientific issues with social, environmental, and political implications, ranging from genetically modified foods to climate change. Understanding and making decisions about these important concerns requires adequate background knowledge and the ability to problem-solve using evidence-based reasoning and critical thinking skills. Teaching these skills—especially scientific and quantitative reasoning—has been the focus of recent national science education reform goals, emphasizing specific strategies to increase scientific and quantitative reasoning (Achieve, 2013a; College Board, 2009; National Research Council [NRC] 2007, 2012). Much has been published on effective science teaching and the assessment of scientific reasoning skills of students (Banilower et al., 2013; Goe, Bell, & Little, 2008; Nicolaidou, Kyza, Terzian, Hadjichambis, & Kafouris, 2011; Sadler, 2004). However, in an age when effective teaching is measured by student achievement on high-stakes tests, determining the alignment between reform goals and assessments is vital.

Similar to classroom reform efforts, the process of evaluating teachers has also evolved. Prior to the 1990s, teachers were evaluated with multiple-choice tests such as the National Teacher Examination (Wilson & Wineberg, 1993). Teacher performance assessments were introduced in the late 1980s, shifting the evaluation of teaching from a quantitative operation to be more in line with professional practice (Darling Hammond, 1986). In the 21st century, standardized tests are again being used for teacher evaluations, but this time teachers are being evaluated on their contribution to their students' performance and overall learning growth measured from high stakes test performance (Colorado Department of Education, 2010; Goe & Holdheide, 2011). During a time when performance on common assessments is carefully analyzed, teachers are often forced to *cover* the easily assessed content of their curriculum in

order to prepare students to be successful on standardized tests. External pressures brought about by policy constrict reform efforts from becoming a reality in our classrooms. This dichotomy between reform and policy plays out in classrooms across the country.

This study examined the extent the tests developed by school district teachers to evaluate their own and their teacher peers' effectiveness were aligned with science education reform goals; specifically, scientific and quantitative reasoning skills. In addition, teachers' ability to develop student assessments was examined. A potential dilemma for science educators is how to develop tests that assess scientific literacy (SL) and use those same assessments to measure their own effectiveness as teachers (Figure 1).

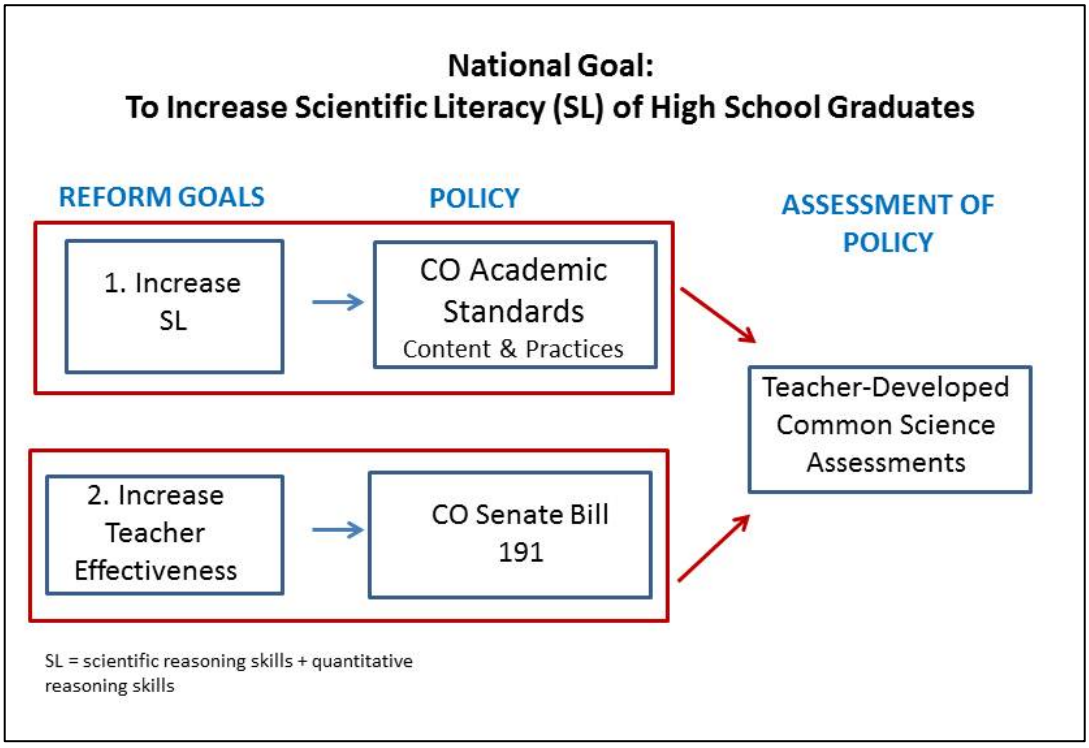


Figure 1. Research focus. This figure illustrates how the two reform goals and the policies enacted to meet them are assessed by the same common science assessment.

Background of the Problem

Scientific literacy reform and policy. The overarching purpose of one reform effort is to increase the SL of high school graduates. The basis for science education reform originates with the ongoing efforts to develop scientifically literate citizens (American Association for the Advancement of Science [AAAS], 1993; NRC 1996, 2012). National science education reform efforts have predominantly focused on increasing SL, science reasoning and 21st century skills, including quantitative reasoning (QR) in science (Duschl, 2012; Duschl, Schweingruber, & Shouse, 2007; NRC, 2012; Organization for Economic and Cooperative Development [OECD], 2009).

SL has been defined in various ways in the literature; since the publication of Taking Science to School (NRC, 2007) the definition has emphasized students' abilities to reason and communicate scientifically and quantitatively. The National Science Education Standards (NSES) (NRC, 1996) highlighted the importance of one's ability to understand 21st century scientific practices and knowledge; the NSES outline the importance of literate citizens' abilities to assess the credibility of information based on its source, methods used to obtain it, and legitimate connection to scientific knowledge.

In 2012, the successor to the 1996 NSES was published. The publication of *The Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas* was the first step in a two-stage process to produce the Next Generation Science Standards ([NGSS], NRC, 2012). The framework, designed to fulfill the long held vision for a scientifically literate public, outlines the essential practices of scientists and engineers. Teachers are expected to imbed the practices throughout their curriculum as they teach the scientific core ideas and cross cutting concepts, instead of what is often performed as separate stand-alone skills. The NRC

(2012) committee calls for the abandonment of the "mile wide, inch-deep" curriculum to achieve a 21st century vision, that emphasizes the development of the practices of scientists and engineers (p.10). An immersion into the practices outlined in the framework ought to be more than just knowledge of content; the aim, is to graduate students with an understanding of what science is, what scientist and engineers do, and how new knowledge is generated.

At the core of SL are the skills of evidence-based reasoning; using scientific knowledge in argumentation and decision-making; and possessing quantitative reasoning skills sufficient to digest multiple lines of evidence (Achieve, 2013a; NRC 2007; 2012). Understanding what science is and the limitations of scientific research also contributes to the SL of citizens (Alchinn, 2011; Lederman, 1992). Evidence-based reasoning cannot be mastered unless one has the ability to make sense of both numerical and observational data. Individuals using quantitative reasoning (QR) are able to gather and collate data, use mathematical equations and models, produce and read graphs, explain the importance of probability, and analyze and interpret the results (NRC, 2007). Moreover, QR skills often overlap with mathematics content standards, which science teachers are now accountable for addressing in the Common Core State Standards initiative (National Governors Association, 2010).

The construct of "scientific literacy" has evolved from a list of concepts gathered in the 1960s to a multi-dimensional framework that includes framing scientific knowledge within the context of history, society, and recognizing the consequences of individual decision-making skills (Trowbridge, Bybee, & Powell, 2008). Organizations such as the National Research Council (NRC, 2007; NRC, 1996; 2012), the American Association for the Advancement of Science (AAAS, 1990; 1993) and the Next Generation Science Standards committee (Achieve, 2013a) all reiterate reform goals related to increasing the level of scientific literacy within the

American citizenry. Common themes are apparent throughout these definitions that were published over a two-decade period (Table 1). Though the variations in the definition of scientific literacy are evident in the science education research literature, for the purpose of this study SL will be defined as: the ability to use scientific knowledge and understanding in the formation of evidence-based scientific explanations, arguments, and personal decision-making.

Table 1

Definitions of Scientific Literacy

Year	Author	Document	Definition
1996	National Research Council	*National Science Education Standards*	The ability to navigate scientific knowledge for personal decision-making and participation in community affairs
1999	Organization for Economic and Cooperative Development	*Program International Science Assessment*	The ability to use scientific knowledge and processes not only to understand the natural world, but also to participate in decisions that affect it
2007	National Research Council	*Taking Science To School*	The ability to understand and evaluate scientific information and to make decisions that incorporate that information appropriately.
2013	Achieve	*The Next Generation Science Standards*	The ability to explain both the natural world and what constitutes the formation of adequate, evidence-based scientific explanations.

An essential element shared by both the common core state standards for English Language Arts (ELA) and the NGSS is to develop persuasive communication skills. In order to achieve this, students must be able to first make sense of evidence/data, then collaborate, engage in discourse, and present findings to a community of peers for critique and debate (Chen, 2011). Assessment frameworks and written argumentation heuristics are sometimes used to scaffold

student's decision-making on various issues (Balgopal & Wallace, 2009; McNeill & Krajcik, 2009; Nicolaidou et al., 2011). However, most studies of scientific communication and discourse are grounded in Toulmin's (1958) argumentation framework and center on oral argumentation in small group or whole class discussions (Acar, Turkmen, & Roychoudhary 2010; Brown, Furtak, Timms, Nagashima, & Wilson, 2010; Erduran, Simon, & Osborne 2004; Evagorou, Jimenez-Aleixandre, & Osborne, 2011; Zohar & Nemet, 2002). In *The Uses of Argument* Toulmin (1958) presented a layout for arguments that begins with a claim supported by data (or evidence). Backing statements are those that are often non-debatable and are used to support the central argument. The type of reasoning, known as warrants in Toulmin's model, is often implicit within the argument. They can be grounded in different types of reasoning, such as rationalistic/ scientific, informal, or moral (Sadler & Zeidler, 2005)

According to both the national and Colorado state standards the skills necessary for developing SL must be fostered in primary grades. In fact, both state and national standards expect students to be able to construct a claim based on first-hand evidence they observe and collect (Colorado Department of Education [CDE], 2010; NRC, 1996, NRC, 2012). Over time, the ability to examine second-hand evidence (gathered by someone else, usually an authority, or scientist outside the classroom) needs to be brought into the curriculum and focused on by teachers. The rationale of the authors of the NSES (NRC, 1996) is that as students graduate to become productive citizens they most likely will not be building models and testing hypothesis first hand. Scientifically literate citizens will need to weed through these second-hand data to identify and evaluate claims, evidence, and reasoning from various sources.

In the 21st century, science, technology, engineering, and mathematics (STEM) permeates nearly every corner of our society. Yet too few Americans have the fundamental knowledge of

them (NRC, 2012). The resultant efforts of reform are in response to the call for a new approach to science education in the United States. The NGSS (Achieve, 2013a) provide an explicit attention to the practices of scientists and engineers. This is different from the NSES (NRC, 1996), which mainly focused on scientific inquiry practices. The NGSS include not only inquiry but also data analysis and interpretation; constructing explanations and arguments from evidence; and quantitative reasoning skills ranging from basic mathematical computations to developing and using models. The NGSS developers are encouraged that the overt focus on practices will result in an increase in SL of students across the United States (Achieve, 2013a).

In spite of the reform goals to increase scientific literacy, American students are still not as scientifically literate as educators and researchers would like (Gess-Newsome, Southerland, Johnston, & Woodbury, 2003; NRC, 2012; OECD, 1999). One explanation for this is the steadily increasing focus on student gains measured by high stakes assessments. Instead of emphasizing the broader application and nature of all subject areas, what is tested has become the driving force behind what is taught in the United States (Johnson, Zhang, & Kahle, 2012).

Assessment of scientific literacy. It is important to define and determine how to assess scientific literacy. Amassing more science content knowledge and random facts about the natural world does not necessarily improve SL skills, according to Anderson (2012), yet this is what is often times the focus of assessments. Henri Poincare the 19th century French physicist stated that, "Science is built up of facts, as a house is with stones. But a collection of facts is no more a science that a heap of stones is a house" (Poincaré, 1917). Even now, a century later, scientists and science educators argue that to be scientifically literate, students must be able to engage in critical thinking and problem-solving activities (NRC, 2012).

In an effort to improve and standardize K-12 educational goals, the Elementary and Secondary Education Act (ESEA) was passed in 1965 as part of President Lyndon B. Johnson's "War on Poverty" (Abott, ND). This national policy attempted to close the achievement gap between low-income and affluent American children. Since 1965, multiple reauthorizations of ESEA have been initiated. For example, No Child Left Behind (NCLB) of 2002 emphasized standardized assessments, local control of schools, and funding tied to accountability (Whilden, 2010). Another outcome of NCLB was the push for all science teachers to have undergraduate degrees in science content areas (US Department of Education, 2006).

Some researchers argue that the emphasis on high stakes assessments in schools has created a *teach-to-the-test* mentality that encourages some teachers to teach to high-stakes tests (Anderson, 2012; Wagner, 2008). Morevoer, as Lawson (201) explains, teachers are concerned that inquiry activities may be too time consuming and take time away from helping prepare students for high stakes assessments (Lawson, 2010). Although frequent testing allows schools, districts, and states to track the achievement and growth of their students, this practice has also been criticized. According to Stiggins (1999), the most detrimental aspect of the testing movement has been that the time spent on test-taking preparation overemphasized basic skills such as content knowledge and neglected higher order thinking skills. Not surprisingly, studies focusing on student achievement in SL and abilities to reason scientifically have shown poor performance (OECD, 2009). Research on the *test prep* curriculum has found that it constrains teacher's use of inquiry practices (Songer, Lee, & Kam, 2002). In order to ensure a scientifically literate society the curriculum and the assessments that measure mastery of it must support one

another. Perhaps developments of quality assessments that measure students' SL skills are needed (Gormally, Brickman, & Lutz, 2012).

Teacher effectiveness reform and policy. Teachers are an important variable and many may be the most important variable that predicts student performance (Dilworth, 2013; Loucks-Horsley & Masumoto, 1999; Parcel & Dufur, 2001). At the same time, Goe et al. (2008) argued that students are not blank slates or passive receptors of information when they arrive in a teacher's classroom and teachers should not treat them as vessels to be filled, nor should they be measuring their progress through multiple choice and fill-in-the-blank tests. This is part due to the fact that there are many additional factors that impact student achievement measures, which include: school context, home and community support, individual student needs and abilities, peer culture and influence, prior teachers and schooling, other current teachers, differential summer learning loss, and finally the specific tests used which rarely measure what they were intended to measure (Darling-Hammond, Amrein-Beardsley, Haertel, & Rothstein, 2012). Although Popham (1999) asserted that educators should be held accountable for student learning, research indicates that other variables should be considered during that process.

Assessment of teacher effectiveness. Much effort has gone into defining what it means to be an effective teacher; notably the *Measures of Effective Teaching* (MET) project (Gates Foundation, 2013) spent 3 years collaborating with 3000 teachers in six urban districts across the country in an effort to create and test measures for effective teaching. The *Framework for Teaching* developed by Danielson (2007) was specifically chosen by the MET project to guide the observational portion of effective teacher evaluations. Four domains guide the framework for teaching including:(1) planning and preparation, (2) the classroom environment, (3) instruction, and (4) professional responsibilities. The framework for teaching does not support a particular

teaching method, nor is it intended to be used as checklist of specific behaviors and assessment strategies, rather effective teachers' classrooms maintain the qualities of cultural sensitivity, high expectations, developmental appropriateness, accommodating individual needs, and appropriate use of technology (Danielson, 2007).

In 2009, another reauthorization of ESEA was enacted; *Race to the Top* legislation created a competitive process during which states could apply for grant funds to enact education reform efforts (US Department of Education [USDOE], 2009b). Race to the Top emphasizes the design and implementation of rigorous standards and high-quality assessments, as well as the attraction and retention of effective teachers and leaders (USDOE, 2009a). As a result of the national competition for funding, a growing number of states and districts are increasing their teacher evaluation systems to include measures of student performance and in an effort to keep pace with educational policy and reform efforts states are exploring which standardized assessments are appropriate to meet their needs (CDE, 2010; Steele, Hamilton, & Stecher, 2010). Colorado legislators were determined to be awarded a Race to the Top grant, thus Colorado Senate Bill 191 (SB191), also known as the *Teacher Effectiveness Bill*, was passed in an effort to be one of the states chosen to receive this funding. School districts can choose to use the model evaluation system developed by the state or create their own system as long as it aligns with state policy. The development of Colorado's evaluation model was guided by key principles including: the use of data to inform decisions; the focus on continuous improvement and feedback; and the involvement of key stakeholders through a collaborative process (CDE, 2010)

Though multiple measures will be used, SB191 mandates that at least 50% of a teacher's evaluation be based on student academic growth. CDE requires that teacher evaluations must include measures of both individual and collective student learning outcomes (SLOs) and when

available statewide summative assessment results and growth measures (CDE, 2010; 2013). Only language arts (3^{rd} –10^{th} grades), mathematics (3^{rd} –10^{th} grades), and science (limited to grades 5, 8, and 10) have assessments in place that can be used to measure students' academic growth on a collective level. As a result, many districts across Colorado are developing district-level common assessments for individual courses to comply with the collectively attributed student learning outcomes requirement of the law (Poudre School District [PSD], 2010; 2013). In a local control state over 175 school districts will be determining what assessments each will enact to comply with SB191 teacher evaluation requirements (CDE, 2012b). The "other 69 %," which refers to the percentage of teachers who teach subjects or grades that are not assessed with standardized tests, do not currently have models to follow that systematically measure student-learning growth. According to some, teachers' jobs have been put on the line, while experimental test development, implementation, and scoring procedures are tried out (Goe et al., 2008; Goe & Holdheide, 2011).

Given the understanding that scientific and quantitative reasoning skills, specifically the use of evidence and reasoning in making claims and decisions are central to demonstrating SL; it makes sense that reform efforts call for these skills to be improved (AAAS, 1990; Achieve, 2013a; NRC, 1996, 2007, 2012; & College Board, 2009). Sadler and Zeidler (2009) believe that it is imperative that science educators assist in preparing the next generation of scientifically literate citizens capable of making meaning of complex socio-scientific issues (SSI). Yet, these ends can only be reached by ensuring that our high-stakes assessments require students' to reason scientifically (Sadler and Zeidler, 2009).

Purpose of the Study

The goals of this study are to determine whether science teachers a) have SL knowledge and skills, b) have knowledge and skills needed to develop district assessments that are aligned with SL practices, and c) develop assessments that are aligned with SL practices outlined in science education reform documents. Even if teachers have knowledge and skills necessary to demonstrate student learning outcomes, they may not choose to do so, knowing that student performance on the instruments may impact their tenure and promotion (Goldhaber & Hansen, 2010). The implications of the findings will inform teacher educators, administrators, and researchers about (1) What professional development teachers might need, and (2) Whether or not asking teachers to develop assessments for their own performance evaluation is problematic or not.

The interrelated constructs that have been developed (reform, policy and assessments (Figure 1) all likely influence an educator's practice relating to SL. Teachers' perceptions of and skills in implementing SL curricula will undoubtedly impact their practice. Concurrently, science education reform practices and initiatives also are likely to impact not only teachers' perceptions and skills within SL, but also their practice (Figure 2).

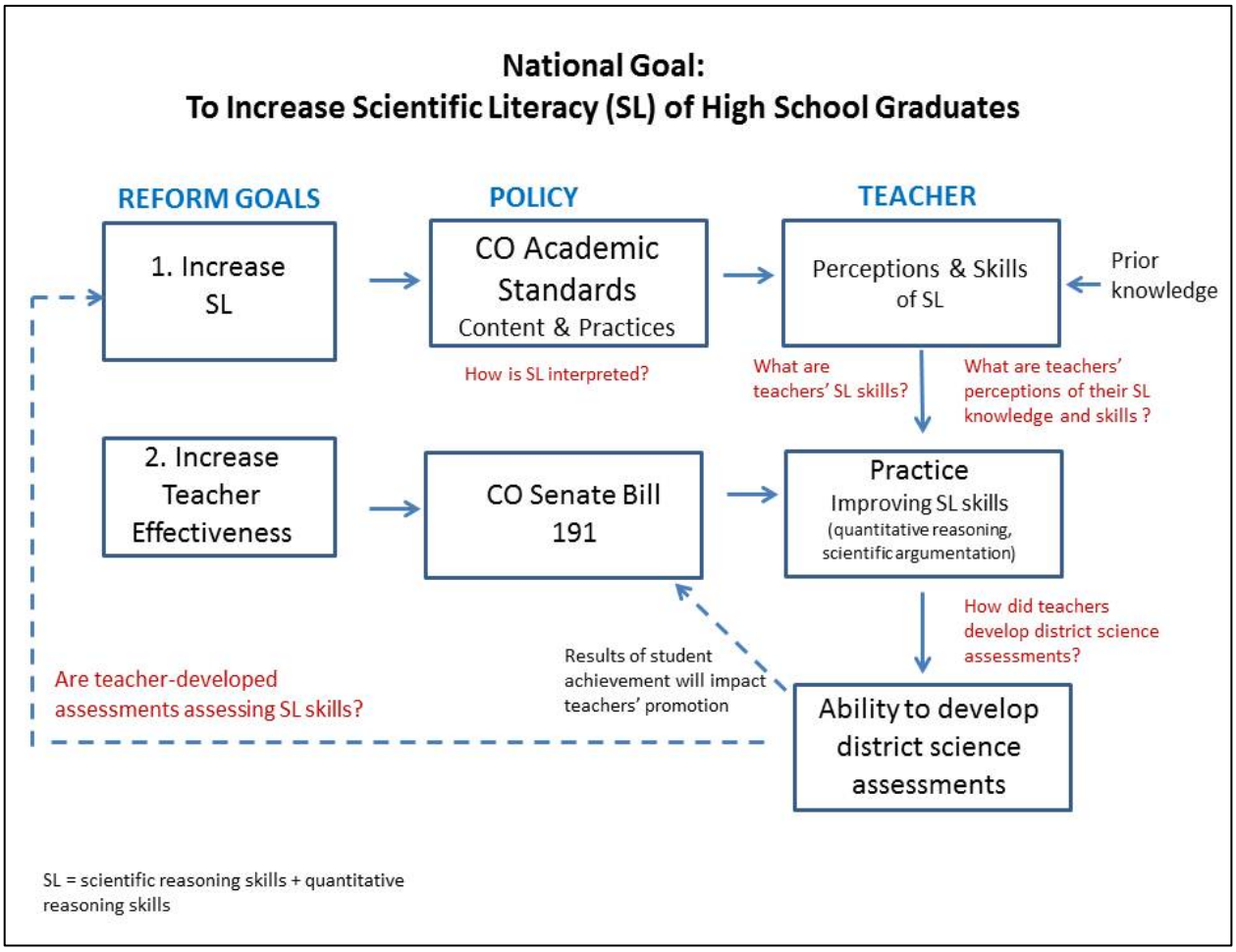


Figure 2. Expanded Conceptual Framework. Teachers' levels of SL, their perceptions of SL, and their self-reported abilities to create assessments that measure SL will be evaluated in this study.

Research Questions

Guided by the expanded conceptual framework, the following research questions directed this study:

1. How do science educators define scientific literacy?
2. What are district secondary science teachers' perceptions of scientific literacy skills?
3. What is the relationship between teachers' overall levels of scientific literacy and the quality of common assessments they create?

Significance of the Study

This study is significant because it contributes to the growing body of research on science education reform efforts and student achievement in science. It will shed light on the dichotomy that teachers face as they strive to provide meaningful learning experiences for their students while at the same time they feel the pressure of policies that impose high-stakes testing systems on the same students. Providing both quantitative and qualitative evidence regarding the extent of scientific reasoning skills in the standards and district developed common assessments to stakeholders can increase their awareness of whether the assessments are aligned with education reform efforts. Specifically, this study is important for district administration, as it will provide concrete feedback on district assessment products, which can inform future professional development planning and policy implementation. Teachers can learn from this study, specifically about the prevalence of scientific reasoning in standards and the need to make room for the teaching, learning, and assessment of these skills in their curriculum. Teacher educators will find this study relevant to their efforts as it may highlight skills that teachers need to improve their understanding of SL and/or assessment development.

Research Design

This multiple methods associative research study includes (a) a content analysis of state and national science standards (middle and high school level); and teacher developed common assessments; (b) a survey of science teachers' knowledge and perceptions of SL as well as their experience with (common) assessment development; and (c) an analysis of teachers' SL levels, using the published Test of Scientific Literacy Skills (TOSLS) (Gormally et al., 2012).

Researcher's Perspective

I have worked in the school district of study for almost 20 years under various capacities, including: Junior High School Science Teacher and Department Chair; District Science and Health Curriculum Facilitator; Teacher-in-Residence (2-year sabbatical at Colorado State University); and most recently high school science teacher. The insight I have gained throughout my tenure assisted me in formulating my research questions. Anticipated limitations of the study include the ability to release my own survey to my peers; having school district administrative assistance in releasing the survey has alleviated this. Additional limitations and delimitations are outlined in chapter three of this study.

Nature of the Study

The overall dissertation consists of five chapters: the introductory chapter; a review of related literature in chapter two; the methodology for the multiple method research study in chapter three; the fourth chapter contains the results and data analysis; and a final chapter summarizing and integrating the results from previous chapters.

In chapter 2, I present the theoretical and conceptual frameworks in which this study is grounded and the background and rationale for the research questions. The review includes a discussion of SL reform goals and policies enacted to measure the achievement of reform.

In chapter 3, I describe the methodology followed for this research study, including the multiple methods used, the instrumentation, sample, and data analysis.

In chapter 4, I present the findings of the study. The quantitative data from the content analysis, the survey data analysis, and statistical analysis correlating teacher survey data and TOSLS scores with content analysis findings are reviewed.

In Chapter 5, I present an overall summary and a conclusion of the study. Recommendations for further research and teacher professional development are made and the implications of the use of teacher developed common assessments as a measure of teacher effectiveness are also posited.

CHAPTER 2: REVIEW OF THE LITERATURE

Classroom teachers have the very challenging job of developing meaningful lessons, differentiating their instruction, and assessing student progress, while managing parent communication, and school, district, and state mandated policies. Further, classroom teachers in Colorado will be annually evaluated in part by the documentation of student growth on state and district common assessments. In some Colorado school districts, secondary science teachers are responsible for developing the district student assessments that will be used to grant tenure and promotion of the teachers. To examine the practice of whether teachers are prepared to develop common assessments it is necessary to first ground this study within a theoretical framework. Subsequently, the conceptual framework (Figure 2) necessitates a review of the literature regarding the nature of SL, SL reform goals, education policy -- including teacher effectiveness initiatives, and assessment related to both reform and SL

Theoretical Framework

Social cognitive theory. An epistemological assumption of this study is that people construct their knowledge through interactions with others in their daily lives; hence, learning is socially constructed (Crotty, 1998). Despite reform efforts aimed at teaching pre- and in-service teachers about learning theories, most science teachers in the United States tend to hold epistemological beliefs aligned with a behaviorist tradition, which presumes that individuals are in control of all of their behaviors (Jones & Carter, 2007). The most influential theory of learning and behavior is the social learning theory (or social cognitive theory; SCT) of Albert Bandura, which emphasizes the reciprocal relationship between environmental and personal factors with behavior (Figure 3, Bandura, 1977). The three factors are constantly affecting each other, for the environment provides models for behavior as people learn through interpreting what they see.

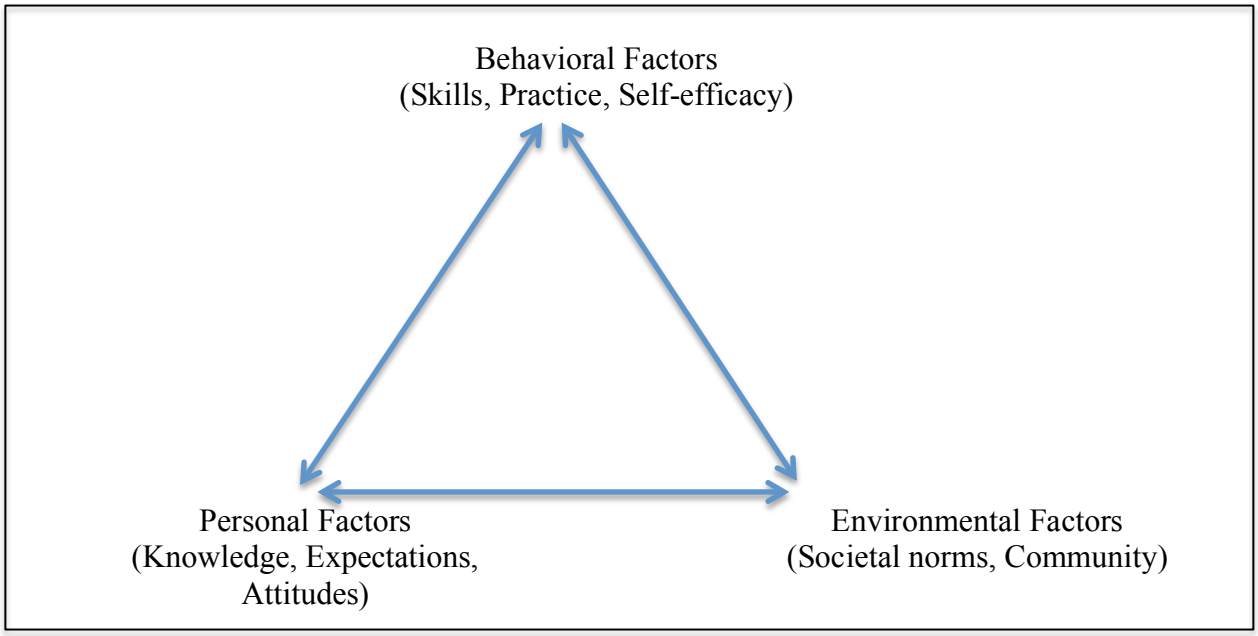


Figure 3: Social Cognitive Theory (Adapted from Bandura, 1977, 1986)

Theories linking beliefs to behavior include the theory of reasoned action, which suggests that a person's actions can be predicted from their intent to perform that action and the theory of planned behavior, which forms the link between beliefs and perceived behavioral control (Zint, 2002). The theory of planned behavior stemmed from self-efficacy theory, which also was proposed by Bandura (1977). Self-efficacy is commonly defined as the belief in one's capabilities to achieve a particular goal or outcome. For teachers to believe that their actions will make a difference Bandura (1986) suggested that teachers need a strong conviction that they can successfully execute a behavior that will produce the desired outcomes in order to act. If any doubts about performance persist, those doubts will drive the behavior.

Drawing on SCT, Jones and Carter (2007) developed the sociocultural model of embedded belief systems, which combines the factors of teacher efficacy, social norms, and environmental constraints, with epistemological beliefs, and instructional practices. The concept of self-efficacy has also been identified in several studies as a major component in the

instructional decision-making of teachers. It has been demonstrated that levels of self-efficacy are indicative of teachers' willingness to implement new instructional strategies or other reform efforts (Jones and Carter, 2007; Zint, 2002).

Pedagogical content knowledge. Beyond self-efficacy, Abell (2007) asserts that teacher behavior is also influenced by his/her pedagogical content knowledge (PCK), or specialized knowledge for teaching, which distinguishes teachers from specialists in the respective field of study. PCK was first described by Shulmann (1986) and refers to the argument that high quality content teachers are able to demonstrate knowledge and practice of both their discipline and of pedagogy. Hence, a pedagogy expert not educated in the natural or physical sciences is less likely to exhibit high PCK within a science classroom. Likewise, a content research expert in the natural or physical sciences may not be able to effectively teach a classroom full of students.

PCK research has been embraced in the science research community since it can enhance science teaching and learning (NRC, 2000a; Magnusson, Krajcik, & Borko, 1999). Park and Oliver (2008) identified the following knowledge components that are included in PCK: (a) orientations toward teaching science, (b) knowledge of student understanding and prior conceptions, (c) the curriculum, (d) knowledge of instructional strategies (e) assessment, (f) school/community, and (g) pedagogy. PCK is a unique domain of content knowledge, that is - it is highly dependent on content that varies from topic to topic or discipline to discipline (Alvarado, 2010; NRC, 2000a; Van Driel, Verloop, & De Vos, 1998), hence, a teacher must able to anticipate specific content misconceptions, how to sequence content instruction, as well as how to assess student knowledge and skills within a discipline.

Mastering PCK in any content area generally requires years of teaching experience (Lee, Brown, Luft, & Roehrig, 2007) and PCK for the science teacher can be tricky since within

science education the content is tightly linked with skills necessary to increase scientific literacy—a construct that is not always easily assessed. Sampson and Blanchard (2012) sought to learn about practicing teachers' knowledge and abilities with scientific argumentation. Through interviews of 30 secondary science teachers they found that teachers relied primarily on their prior content knowledge rather than available data when making arguments, they also found that teachers rarely used support in the form of data and reasoning in their arguments (Samspon & Blanchard, 2012). Crippen's (2012) case study involving 42 high school teachers had similar findings. Teachers took part in a two-week summer institute focused on developing content knowledge and argumentation skills related to climate change. Despite the volume of evidence provided first-hand to the teachers, the majority of argumentative evidence teachers used came from outside resources (Crippen, 2012).

McNeill and Knight (2013) studied outcomes of teacher professional development aimed specifically at increasing teachers' argumentation PCK. The claims, evidence, reasoning framework was taught during the workshop and the greatest gains between pre and post assessments were in the teachers' ability to discuss students' reasoning in their written arguments. Yet they noted that teachers had more difficulty applying their new knowledge to create activities related to argumentation. With the broad support of PCK, and the commonality of scientific literacy (science and quantitative reasoning; scientific argumentation) as content across all grade levels, the research on science teacher's PCK of argumentation is continuing to grow.

Conceptual Overview of Scientific Literacy

Examining how or what it means to reason scientifically requires an understanding of: the historical perspective of science education reform and scientific literacy, which includes the

definitions of science, NOS, and scientific inquiry. In addition, understanding what it means to become scientifically literate in the 21st century requires research into evidence-based scientific reasoning, decision-making, and argumentation. Teachers' efforts to improve students' SL, however, are impacted by science education reform policy. The focus on the current state of science education reform, specifically as it relates to effective science teaching and developing quality assessments aligned to science education reform goals, will be studied.

If one were to ask a group of people to answer the question: "Without listing any scientific disciplines, what is science?" the responses might be as unique as each individual. While the purpose of this review is not to delve into the history of scientific thought, discovery, and theoretical underpinnings, it is important to ground the review in an operational definition of science and how it is situated in the evolution of the definition of SL. The discipline of science is unique because it represents both a way of knowing and the knowledge generated from this methodology. Scientific knowledge is collectively made up of the facts, the theories (explanations of natural phenomena), and the laws of science (descriptions of natural phenomena). Although the public may believe that scientific facts are the most important aspect of science, scientists are more likely to argue that scientific theories are central to the discipline (Scott, 2004). It is important to note the tentative nature of knowledge, which can arise from the discoveries and re-verification of new evidence (Kuhn, 1962). The ways of knowing in science include (but are not limited to) the accepted empirical methods of inquiry, research driven by testable hypothesis, repeatability and falsification of evidence, data driven reasoning, and peer review (Understanding Science, 2012). Most often, scientific research involves deductive reasoning; however, there are many examples of new science knowledge being generated

through inductive studies. Inductive studies, such as the documentation of new genes or species are equally important contributions to scientific knowledge, but are often followed by deductive studies that examine more specific features or phenomena associated with these discoveries. Hence, scientific knowledge expands, as scientists are able to generate new knowledge through the advent of new technology or through the repeated testing that scientific ways of knowing demand.

Historical perspective. Science education reform spans numerous decades beginning in the 1950's through present day. The national (American) call to improve scientific literacy began in 1957 when the first artificial satellite, Sputnik, was launched by the Soviet Union (Cadbury, 2006). This was the beginning of what was known as the *space race*, which United States President Dwight Eisenhower decreed was necessary to safeguard our country from space-based missile attacks (Divine, 1993). It has since been referred to as the "Sputnik moment" by the media, policy makers, science education reformers, and even the president of the United States (Harmon, 2011). Americans were concerned that other nations were investing more resources in science and mathematics education in grades K-12 than we were. In July of 1958, six months after the launch of the first American satellite – Explorer I, Eisenhower established the National Aeronautic and Space Administration (NASA). The post-Sputnik era produced numerous science curriculum projects, including efforts to humanize science education through the use of the history of science (Wang and Marsh, 2002). American students benefitted from these programs designed to improve math and science education nationwide. In addition to the many innovations in education and research, the space race launched a technological revolution and many innovations produced for space travel have been adapted for non-aerospace uses by the private sector (Garber, 2007).

The push for a scientifically literate society followed the launch of Sputnik. At that time Snow (1959) identified the divide between intellectuals and scientists, which in his view represented a "gulf of mutual incomprehension, hostility and dislike, and most importantly, resulted in a lack of understanding" between the two worlds (in Laugksch, 2000, p. 76). In 1966, Pella, O'Hearn, and Gale undertook an empirical review of 20 years of SL definitions. They concluded that an individual who was scientifically literate understood (a) interrelationship between science and society, (b) ethics that control the scientist in his work, (c) the nature of the scientific process, (e) difference between science and technology, (d) basic concepts in science, and (f) interrelationships of science and the humanities (Pella et al., 1966).

Showalter (1974) compiled 15 years of research into seven dimensions of SL which at the time were deemed to have a high degree of specificity, yet continued to reiterate the same conceptual understandings (e.g., understanding NOS; application of science concepts, laws, and theories; using science to solve problems). Following Showalter, Shen (1975) suggested three components of SL: practical, civic and cultural. His suggestions were the first to involve the citizenship component of SL, which aimed to enable citizens to use science in decision-making related to societal issues.

Into the 1980s, definitions of SL began to portray the importance of SL in a democratic society (Arons, 1983). Scientists and science educators, in conjunction, pushed for 'Science for all' in the 1990's and advocated for curricula that were relevant and applicable to all students (AAAS, 1989). *Science for All Americans* ([SFAA] AAAS, 1989), acknowledged the need for SL beyond individual self-fulfillment and even immediate national issues by identifying a rather long and alarming list of the serious global problems facing our planet (e.g. population growth, shrinking rainforest and species diversity; pollution; extreme inequities of earth's wealth and

resource consumption). SFAA identified their multi-faceted definition of SL as one that encompasses mathematics and technology in addition to the natural and social sciences including:

> Being familiar with the natural world and respecting its unity; being aware of some of the important ways in which mathematics, technology, and the sciences depend upon one another; understanding some of the key concepts and principles of science; having a capacity for scientific ways of thinking; knowing that science, mathematics, and technology are human enterprises, and knowing what that implies about their strengths and limitations; and being able to use scientific knowledge and ways of thinking for personal and social purposes. (AAAS, 1989, p. 20)

The 'Science for all' era was followed by the standards-based science education reform movement, which included the publication of the NSES in 1996. These national standards offered a vision for what it means to be scientifically literate and described what every high school graduate should know and be able to do as a result of their education (Bybee, Powell, and Trowbridge, 2008). The NRC (1996) highlighted the importance of one's ability to navigate scientific knowledge for personal decision-making and participation in community affairs. This stance included the understanding of what science is and how new knowledge is generated through scientific inquiry, as well as the importance of being an informed consumer of that knowledge.

The nature of science (NOS). Science as a way of knowing, or the epistemology and sociology of science, can be referred to as NOS, which some science education researchers describe as the values and beliefs inherent in the development of scientific knowledge (Abd-El-Khalick & Lederman, 2000; Lederman, 1992). According to Bell, Abd-El-Khalick, and

Lederman (2000) public understanding of NOS is a critical component of democracy, in which people must make decisions on science and technological issues. It can be concluded that an adequate understanding of NOS is central to scientific literacy (Abd-El-Khalick & Lederman, 2000, p. 665). Yet, often individuals conflate NOS with science process skills, which can be confusing (Lederman, Abd-El-Khalick, & Schwartz, 2002). For example, the skills of observation and abilities to hypothesize are considered science process skills, whereas NOS constructs are guided by the understanding that observations can be enhanced or constrained by the equipment available, and hypotheses can require imagination and creativity (Lederman et al., 2002).

NOS can also be referred to as the *work scientists do* (NRC, 2012). This includes the continual testing of hypotheses that can support scientific theories – or not. Science does not *prove* anything; rather, scientists disprove alternative hypotheses. If the evidence supports a scientific proposition, and if the scientific community accepts the scientific and quantitative reasoning employed in the scientific argument, this proposition is accepted until other evidence disproves it (AAAS, 1990). Understanding the iterative and dynamic nature of scientific inquiry is at the heart of being scientifically literate.

Scientific inquiry. A scientifically literate citizen must possess an understanding of scientific inquiry, which often begins with observations of the natural world. Inquiry, as used by teachers in a broad sense, refers to instruction that supports students questioning about content. It is often facilitated by teachers questions (Bybee et al., 2008). Scientific inquiry, therefore, is the questioning that accompanies scientific study in an effort to better understand scientific processes (Lawson, 2010). Scientific inquiry is an essential component of scientific literacy and it is thoughtfully highlighted in both state and national science content standards. The NRC

(2000b) publication, *Inquiry and the National Science Education Standards*, identifies the essential features of classroom inquiry, which includes engagement in scientifically oriented questions, giving priority to evidence, formulating, communicating, and evaluating explanations. The applicable skills of scientific inquiry include asking testable questions, researching, predicting, planning, data gathering, analyzing, and making conclusions based on evidence. Additionally, inquiry instruction must include not only the features of inquiry, but also the understandings of inquiry. Teachers must emphasize the beginning of the definition, which puts the focus on the work scientists do to fulfill the requirements of teaching all the standards (NRC, 1996). Whether one uses the moniker *understandings of science inquiry*, *second-hand* inquiry, NOS, or *the work scientists do*, it is imperative that all students are proficient in these understandings. Students must understand how scientists work –- from asking and researching testable questions, to completing controlled experiments, to ultimately analyzing and interpreting data and communicating their evidence-based conclusions.

Understanding inquiry and the work scientists do is one component of scientific literacy that was prevalent in science education reform literature of the 1990's, yet to ensure more was being done to improve the scientific literacy of United States citizens, the update of the National Science Standards in the 21st century was completed in 2013. Advancements in cognitive learning theory (NRC, 2000a) and a better understanding of the way science functions guided the development of the Next Generation Science Standards (NGSS). The introduction of the term practices in the 2012 Framework for K-12 education (NRC) alarmed some educators, yet upon further evaluation of this new perspective one could see that the framework did not replace inquiry with practice; instead it expands and enriches the science classroom with an emphasis on engagement in the practices of scientists and engineers. The NGSS calls for practices in addition

to those that are inquiry based, these include: (1) developing and using models; (2) analyzing and interpreting data (including graphical and statistical analysis); (3) Constructing explanations and designing solutions; and (4) engaging in argument from evidence (Achieve, 2013a; Bybee, 2011; NRC, 2012). The practices are the new basic in science education and the change from inquiry to practice, though most likely a challenge will enhance the scientific and quantitative reasoning skills of all students. Even though the concept of scientific literacy has been promoted since the 1960's, Demir and Abell (2010) found in their review and study of teacher perceptions that there is still a wide range of interpretations of what scientific inquiry, as well as what scientific literacy entail.

Contemporary Views of Scientific Literacy

The dichotomy between content and pedagogy in science education reform and the pursuit of SL continued into the 1990's. Shamos (1995) identified his three levels of SL beginning with "cultural." In contrast to Shen's (1975) cultural component of SL, Shamos identified this level as the basic science knowledge one is expected to obtain during schooling. Shamos' second level or "functional" SL is characterized by the ability to converse, read, and write coherently in a non-technical but meaningful context. Shamos' highest level or "true" SL involves an "individual who actually knows something about the overall scientific enterprise" (p. 89). Shamos (1995) cited John Dewey's century old *habits of mind* philosophy that challenged schools to allow students to learn how to learn from the experiences of life itself.

Laugksch (2000) proposed a classification scheme of the concept of SL according to three different interpretations and uses of the word literate: "literate as learned; literate as competent and literate as able to function minimally in society" (p. 82). The first level—literate as learned—is basically the intellectual value of being scientifically literate, it is based on an

existing body of knowledge to be mastered. The second category – literate as competent, moves SL up to an operational level, requiring the individual to interact and perform activities such as solve problems, think critically and "deal sensibly with problems involving evidence, quantitative considerations, and logical arguments" (p. 83). Category three requires the individual to use science in performing a function in society within their various roles as consumer as citizens (Laugksch, 2000). Being scientifically literate, therefore, involves both knowing science and how new knowledge is generated, and knowing how to communicate this knowledge (Norris & Phillips, 2003).

Norris and Phillips (2003) challenged the previous decades' focus on SL as content or SL as necessary for societal functioning and returned to the fact that "literacy in the fundamental sense is central to scientific literacy" (p. 237). Similar to Shamos' definition of functional SL, fundamental literacy is the constructive process of inferring meaning from text, some of which requires constructs of understanding beyond scientific knowledge (Norris & Phillips, 2003). This may seem like a simple view of SL, but without reading and writing, science could not even be possible. Roberts (2007), in a chapter entitled, *Scientific Literacy/Science Literacy* (abbreviated together as SL) in the *Handbook of Research on Science Education,* (2007), concluded that there is no known consensus of the definition of SL, with one exception that everyone agrees to: students cannot be scientifically literate without knowledge of some subject matter. Citing that SL was originally introduced to professional science educators as an "educational slogan" – or a "way to rally support for reexamining the purpose of school science" (p. 736), it began the discourse into what students, who may not be potential scientists, should know and be able to do. Roberts' research of over 30 years' worth of SL definitions informed his identification of two *visions* of SL (Vision I & Vision II), which he stated represented the extremes between science

subject matter and situations where that subject matter can and should play a role in human endeavors. Vision I, according to Roberts (2007), is rooted in the historical products and processes of science and has been the starting point for SL definitions, whereas Vision II relates SL to matters beyond school and into situations that students are likely to encounter as citizens.

Allchin (2011) tied together the concrete value of what could be considered Vision II SL for preparing citizens. He stressed that for development of scientific literacy students should know what science is, how it works and be able to interpret the reliability of scientific claims in personal and public decision-making. Scientific research findings, in general, will be kept in check more often, if more people routinely examined the credibility of evidence, understood NOS, and were overall more scientifically literate (Allchin, 2011).

Contemporary reform efforts continue to push for scientific literacy. Understanding inquiry and the work scientists do is one component of scientific literacy that was prevalent in science education reform literature of the 1990's, yet to ensure more was being done to improve the scientific literacy of United States citizens, the update of the NSES in the 21st century was undertaken by the NRC. In the National Research Council's critical synthesis of the latest educational research for K-8 science teaching: *Taking science to school: Learning and Teaching Science in Grades K-8* (NRC, 2007), clear emphasis is placed on the understandings of scientific inquiry. Four major scientific proficiencies are outlined for teachers to emphasize:

1. Know and use scientific ideas
2. Generate and evaluate scientific evidence and explanations
3. Understand the nature and development of scientific knowledge
4. Participate productively in scientific practices and discourse

Three of the four proficiencies focus on the abilities and understandings of scientific inquiry and NOS, these outweigh the single proficiency focused on knowing and using scientific ideas.

The *Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas* (NRC, 2012) highlights the most current research in science education. Advancements in cognitive learning theory (NRC, 2000a) and a better understanding of the way science functions guided the development this document, which would form the foundation for the writing of the Next Generation Science Standards (NGSS) in 2013. The committee for the conceptual framework recommended that a coherent approach to K-12 science education, consisting of three dimensions (Scientific and Engineering Practices; Crosscutting Concepts; and Disciplinary Core Ideas) would provide the foundation for students to achieve goals of SL by the end of 12th grade.

The eight scientific and engineering practices describing what students should know and be able to do are modeled after the four proficiencies outlined in the mentioned Taking Science to School (NRC, 2007) publication. NRC's (2012) focus on the use of the term practice(s), rather than standards or proficiencies, reinforces the need for science to be taught through the integration of both knowledge and skills at the same time. In addition, the focus on practices aims to alleviate the often over accentuated single, *scientific method*, which is prominent in many science classrooms (NRC, 2012). The eight practices highlighted below identify the essential scientific and quantitative reasoning skills for K-12 science and engineering classrooms in the 21st century.

1. Asking questions (for science) and defining problems (for engineering)
2. Developing and using models
3. Planning and carrying out investigations
4. Analyzing and interpreting data

5. Using mathematics and computational thinking
6. Constructing explanations (for science) and designing solutions (for engineering)
7. Engaging in argument from evidence
8. Obtaining, evaluating and communicating information (NRC, 2012, p. 49)

The introduction of the term practices in the 2012 Framework for K-12 education (NRC) alarmed some educators, yet upon further evaluation of this new perspective one could see that the framework did not replace scientific inquiry with practice; instead it expanded and enriched the science classroom with an emphasis on engagement in the practices of scientists and engineers. These new standards call for practices in addition to those that are inquiry based, while assisting teachers with the differentiation between how scientists and engineers engage in the practices (Achieve, 2013a; Bybee, 2011; NRC, 2012). The NGSSs (Achieve, 2013a).

The NGSS, modeled after the work of NRC (2012), were introduced to help students understand how knowledge is produced, giving them direct experience with the variety of ways questions are investigated, findings are modeled and explanations of the natural world are based on evidence (Achieve, 2013a). The practices are the new basic in science education and the change from inquiry to practice, though most likely a challenge will enhance the scientific and quantitative reasoning skills of all students.

The definitions for scientific literacy presented earlier were summed up as the ability to use scientific knowledge and understanding in the formation of evidence-based scientific explanations, arguments and personal decision-making. More simply, scientific literacy is *scientifically informed decision-making*. In the 21st century, as society witnesses the plethora of scientific concepts intersecting with human culture it can no longer be up to scientists to *fix* the problem, rather it is going to take a community of scientifically literate citizens to make

scientifically informed decisions to solve the issues facing society today. Ultimately, environmental and socio-scientific problems will be tackled with a common understanding of both the benefits and limitations of science.

Scientific argumentation. Scientific argumentation and decision-making are fundamental to all of science and must be at the forefront of science classrooms (CDE, 2009; Driver, Newton & Osborne, 2000). Research studies have identified student difficulties with evaluation of evidence, socio-scientific argumentation, and decision-making beginning in elementary school and extending through college students (Acar et al., 2010; Nicolaidou et al., 2011; Sadler, 2004). An evaluation of students' decision-making abilities by Sadler and Zeidler (2005) indicated that science teachers should promote other type of non-scientific reasoning (intuitive and moral), in addition to scientific (rationalistic) reasoning. They argue that all three of these types of reasoning are involved in every-day evidence-based decision-making. Although teachers must help their students distinguish between every day and scientific decision-making. Toulmin's (1958) argumentation framework has been well accepted in the science education community (Acar et al., 2010). The credibility of evidence and written argument heuristics are common means used to scaffold student's decision-making on various issues (McNeill & Krajcik, 2009; Nicolaidou et al., 2011). Toulmin (1958) devised a practical framework that is still in use today (Figure 4). An argument can begin with a claim (C; the conclusion or assertion), but without data (D; evidence) to back it up, it will easily be challenged and dismissed. Thus, what follows is the introduction of implicit warrants (W), which can be distinguished from the explicit nature of data (D). Further qualifiers (Q) are necessary which support the strength of the warrant, and warrants also need backing (B) which provide the explicit justification for the claim. It is within the backing that scientific knowledge is crucial, for it provides the

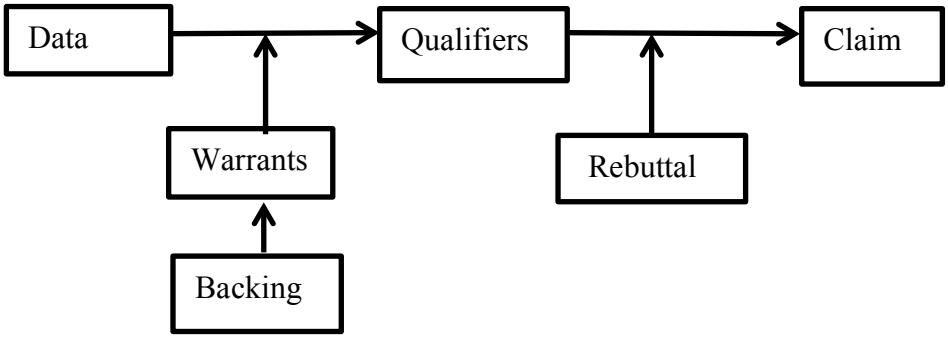


Figure 4. Toulmin's (1958) Framework for Argumentation. This figure highlights Toulmin's model for argumentation adapted for use by science educators.

authority that can be directly linked to conclusions with supporting scientific ideas. Lastly, the most sophisticated argument will have a rebuttal (R), which indicates those non-arguments that illustrate the flaws of undeveloped arguments. Rebuttals in scientific argumentation are statements that when supported by counter-evidence provides additional rationale for the claim (Erduran et al., 2004).

Evagorou, Jimenez-Aleixandre, and Osborne (2012) used the Toulmin framework in a high school class in England to examine how students from different socio-economic and ethnic backgrounds wrote and discussed arguments about the SSI about the control of introduced squirrel species that were outcompeting native squirrels. The researchers found that explicit instruction on argumentation helped students construct more robust arguments about this issue (Evagorou et al., 2012). In a similar study, Evagorou and Dillon (2012) uncovered that class discussion characterized by the teachers' constantly questioning what students are saying (not imposing knowledge on them) resulted in more students' improvement of their argumentation skills.

Within Toulmin's argumentation layout, the greatest potential for error when considering scientific argumentation is between the warrants and backing, since warrants can be implicit or informal. According to Sadler and Zeidler (2005) students informal reasoning often comes into play during argumentation efforts involving socioscientific issues (SSIs). Sadler (2004) reviewed 13 articles to determine the difficulties of students' reasoning while trying to understand SSI's. Three key conclusions were extracted from Sadler's study.

1. Students do not commonly use scientific evidence to support their personal decision making.
2. Students are not competent at analyzing and evaluating arguments.
3. Students often make unjustified claims and struggle to recognize opposing arguments.

Similar to Sadler's review, Sampson and Clark's (2009) study involving 168 high school chemistry students determined that when students engage in argumentation they are challenged with generating a coherent claim, using sufficient and appropriate data to support the claim and understanding what counts as acceptable evidence. Sampson and Clark (2009) further investigated whether collaborative groups craft better arguments than individuals and they concluded that collaboration did not improve the quality of a groups' argument.

Kolsto's study involving 22 students in Norway (2006) identified five types of students' arguments: the relative risk argument, the precautionary argument, the uncertainty argument, the small risk argument, and the pros and cons argument. Not surprisingly, within these arguments students made use of a range of both scientific and non-scientific knowledge. Empirical results from Nicolaidou et al. (2011) found statistically significant differences between pre and post-test scores in the areas of *Students Conceptual Understanding* and *Students Capacity to Assess the Credibility of Evidence.* Results by Venville and Dawson (2010) included an overall shift from

informal (intuitive and emotive) reasoning prior to argumentation instruction centered on an SSI to more formal (rational) reasoning on post assessments.

The focus on argumentation in recent science education reform documents (NRC, 2007; 2012) requires a shift in both the desired student learning goals, and the teachers' role in the classroom (McNeill & Knight, 2013). *Taking Science to School* (NRC, 2007) stress the importance of students' participation in scientific discourse and that teachers must help students grasp the difference between scientific arguments, which are grounded in plausibility and evidence, and everyday arguments, which usually rely on power and persuasion. The seventh practice of scientist and engineers (NRC, 2012) specifically calls for students to engage in argument from evidence and by twelfth grade students should be able to evaluate both their own arguments and those related to controversies in the development of scientific ideas and socio scientific issues.

Argumentation is not only viewed as a goal for scientific literacy, but the skills of writing evidence-based arguments can be found in the Common Core State Standards for English Language Arts (CCSS - ELA) as well. For example, this writing standard for grades 9-10: "Write arguments to support claims in an analysis of substantive topics or texts, using valid reasoning and relevant and sufficient evidence," (Common Core State Standards Initiative [CCSSI], 2013a, par.3) mimics the Science and Engineering Practice *Engaging in Argument from Evidence*, which requires development of a students ability to use "appropriate and sufficient evidence and scientific reasoning to defend and critique claims and explanations about the natural world (s)" (Achieve, 2013b, p. 13). Adopted in 45 states, the CCSS's can support the goals of developing scientifically literate citizens (CCSSI, 2013b).

Evidence-based scientific reasoning. It seems odd that according to Google, almost no one searches for the phrase 'evidence-based reasoning' (73 times/month vs. over 300,000 times/month for "critical thinking"). Yet, identifying evidence and drawing logical conclusions based on this evidence lie at the heart of problem solving, learning, and, yes, critical thinking (Midgley, 2009). Beginning in the primary grades both state and national science standards expect students to be able to make a claim based on first hand evidence they observe and collect (CDE, 2010; NRC, 1996). Over time, the abilities to examine second-hand evidence (that performed by someone else, usually a professional scientist outside the classroom) become an even greater skill, since we know that as students become productive citizens they most likely will not be building models and testing hypothesis first hand (OECD, 2009). Scientifically literate citizens must possess the critical skills to weed through this second-hand data to identify claims, evidence, and reasoning from various sources.

Evidence-based reasoning (EBR) has been a science education reform goal for over 50 years. (Brown et al., 2010). The process outlined in Brown et al.'s (2010) EBR framework is a combination of Toulmin's (1958) framework of argumentation with Duschl's (2003) framework of scientific inquiry; it necessitates analysis, interpretation and application of evidence to produce a claim. Students must also learn that evidence or data cannot and should not be taken at face value. Nicolaidou et al. (2011) developed a credibility assessment framework (CAF) to support students' examination of evidence at a deep level. The CAF interprets credibility as the determination of confidence levels in evidence through the use of formal questioning techniques. Determining credibility requires consumers to critically evaluate the evidence through consideration of critical questions such as, what is the origin of the evidence? Or what is the authority of the scientific source? The results of Nicolaidou et al.'s (2011) study involving the

use of their CAF indicated positive outcomes, yet their research literature paints a different picture (Table 2). Each study examined students' abilities to evaluate the credibility of evidence and to construct evidence-based explanations at various levels. The conclusions of all these studies were the same -- students have difficulty demonstrating these skills.

Table 2

Summary of Studies Examining Students' Evidence-based Reasoning Skills (from Nicolaidou et al. (2011)

School Level	Research Study Author(s)	Findings
Elementary	Sandoval & Cam, 2011; Wu & Hsieh, 2006	Students lacked the criteria they needed to evaluate evidence.
Middle School	Chinn, Duschl, Duncan, Buckland, & Pluta, 2008; Glassner, Weinstock, & Neuman, 2005; Kyza & Edelson, 2005; Mathews, Holden, Jan, & Martin, 2008; Yoon, 2008.	Students had difficulty with effective use of reasons and evidence, and with understanding the need to provide more elaborate justifications.
High School	Brem, et al., 2001; Dawson & Venville, 2009; Reiser, 2004; Sandoval & Millwood, 2005	Students were uncertain of the trustworthiness of second hand arguments and opinions related to SSIs.
Undergraduate	Halverson, Siegel, & Freyermuth, 2010; Korpan, Bisanz, Bisanz & Henderson, 1997; Lippman, Amurao, & Pellegrino, 2008; Treise, Walsh-Childers, Weigold, & Friedman, 2003	Students were inconsistent in identification of credible evidence and tended to select the 'easier' to comprehend choice of evidence rather than the importance of empirical evidence and rebuttals.
Pre-service	Zembal-Saul, Munsford, Crawford, Friedrichsen, & Land, 2002	Students still exhibited limitations grounding their arguments with evidence as reported in the literature.

Both the EBR and CAF frameworks are useful tools for science teachers to use in their classrooms in order to reach the goals of scientifically literate high school graduates. Yet, another aspect of science education reform impacts what goes on in classrooms across the country – that of education policy related to student achievement on high stakes assessments.

Education Reform Policy

The same policies that direct states to develop quality standards and/or adopt the new common core standards also include a requirement of assessing student achievement of these standards.

No Child Left Behind. Examining education policy that attempts to improve student achievement dates back numerous generations with the original Elementary and Secondary Education Act passed in 1965. In 2001 NCLB legislation a reauthorization of the elementary and secondary act was passed. NCLB sought to close the achievement gap by ensuring that high quality assessments, accountability systems preparing and training teachers, ensuring they were highly qualified in the area they taught (US Department of Education, 2001). Measuring student achievement gains by high stakes testing, while part of the education system for over a century, took on a new level of priority with the passage of this legislation. NCLB was a pivotal point in the measure of school accountability, and even though the actual policy contained allocations for improving achievement in students from all subgroups, the bottom line ended up being the percent of student proficiency in reading, writing, and mathematics.

Race to the Top. Current policy continues to rely on student achievement data and has shifted into the realm of evaluating teachers based on student achievement (U.S. Department of Education, 2009b). Race to the Top is a competitive grant program established by President Barack Obama's American Recovery and Reinvestment act of 2009. The four main education reform areas of Race to the Top include:

- Adopting common standards and assessments aimed at preparing students to be post-secondary ready;

- Building data systems that measure student growth and inform teachers and principals about how they can improve instruction;
- Recruiting, developing, rewarding, and retaining effective teachers and principals; and
- Turning around our lowest achieving schools. (U.S. Department of Education, 2009a)

While standards, have had a longstanding history in the United States, dating back over 100 years when the Committee of Ten outlined standard curricula for secondary schools (Committee of Ten, 1894), many consider the development of the common core standards for English Language Arts and Mathematics as just the beginning of a possible movement towards a national curriculum (Carr, Bennett, & Strobel, 2012).

Developers of the CCSS collaborated with the partnership for 21st Century Skills (2002), which resulted in the inclusion of these skills in the new standards. Following in the footsteps of common core are the policies, which hold teachers accountable for their students' achievement scores. The NGSS, as stated previously, were based on the foundation for scientific literacy outlined in the K-12 Science Education Framework (NRC, 2012). The *Framework* principally concerns itself with what all students should know in preparation for their future in the 21st century; furthermore, it integrates science *and* engineering practices, raising engineering design to the same level as scientific inquiry (Achieve, 2013a). The NGSS's incorporate the vision of the *Framework*, calling for the practices of science and engineering to be taught in context – not as a separate entity (Achieve, 2013a). The intention of the science and engineering practices outlined in the NGSS is not to provide specific teaching strategies or curricula, but achievement indicators and learning goals to guide reform-based instruction in K-12 science classrooms across the country.

Stakeholders working on the Next Generation Science Standards (Achieve, 2013a) used the practices outlined in the *Framework* as the specific parameters for writing each grade level performance standard. This guideline is significant since it provided a different approach to standards that was not seen before at the national level. Achieve foreshadowed that

> In the future, science assessments will not assess students' understanding of core ideas separately from their abilities to use the practices of science and engineering. They will be assessed together, showing students no only "know" science concepts, but also, students can use their understanding to investigate the natural world through the practices of science inquiry, or solve meaningful problems through the practices of engineering design (Achieve, 2013b, p. 1).

Time will tell whether this prediction about a change in assessment will become reality for K-12 education in the United States.

Teacher effectiveness measures. In addition to teaching the science content standards (life, physical and earth), teachers must focus their instruction on important process skills, as well as the understandings of scientific inquiry. The National Science Education Standards (NRC, 1996) expects teachers to have both theoretical and practical knowledge related to science teaching and learning; this includes specific skills related to scientific literacy. These practices are the: (a) incorporation of secondary sources and literature reviews; (b) promotion of scientific discourse (use of scientific language) in the classroom; (c) provision of appropriate and ample time for sense making during instruction; and (d) encouragement of curiosity and inquiry by challenging students with focused scientific questions. To increase students' scientific literacy, science teachers must use the process skills of scientists to teach science content, while

reinforcing the practice of the scientist, which includes collaboration, peer review, the use of models, and repetition/replication of investigations.

The State of Colorado updated their model content standards in 2009 and within each of the content areas one can find the emphasis on important 21st Century Skills, as well as inquiry, relevancy, and the *nature* of every discipline (CDE, 2009). The science standards were reduced to three content standards (Physical, Life, and Earth Systems) with abilities of inquiry and NOS embedded within. The revision process was guided by the state's outline of a number of prepared graduate (readiness) competencies, which identify the important aspects of the nature and application of science. The explicit use of crosscutting themes in the standards is included with the expectation that all Colorado graduates will be equipped to succeed in their future (CDE, 2009).

The competitive nature of Race to the Top funds pushed states to implement reform efforts that would further assist them in receiving large sums of grant monies. The state of Colorado was already well prepared in the standards category In addition; Colorado had developed a tool that analyzed student growth from one year to the next. This growth model enabled schools and teachers to analyze the annual growth levels of their students on the Colorado Student Assessment Program or CSAP test, ensuring that students were achieving *adequate yearly progress* on the standards (CDE, 2009).

To fill the gap on measuring teacher and principal effectiveness and thus increasing their chances for Race to the Top funding, Colorado Senate Bill 191 (SB191), known as the "Teacher Effectiveness Bill," was signed into law in May of 2010. SB191 mandates that at least 50% of teachers' evaluations be based on student academic growth (CDE, 2010). Therein lies the problem; only language arts and mathematics have state-level common assessments in place that

could be used to measure student's academic growth at a collective level. As a result of SB191, many districts across the state are developing common assessments in each of the subjects that do not have state-level tests. These assessments can be used for individual teacher evaluations as part of a *body of evidence* measuring student growth.

Education reform policy such as the adoption of the common core standards by states (US Department of Education, 2009a) and the teacher effectiveness measures (CDE, 2010) have the potential to improve our schools and increase levels of student achievement. Yet like so many of the previous efforts to increase student outcomes, the focus on testing may outweigh the original visions for reform (Anderson, 2012).

Assessment in Education

Ever since the early 1900s when Thorndike developed hundreds of achievement tests that were widely implemented (Shepard, 2007), accountability in the form of standardized tests measuring student achievement created a predominant focus on assessment in schools. Tests scores are used for a variety of purposes, to provide feedback to students and their teachers about learning progress, to inform those outside the classroom of the overall academic achievement, and for possibly placement or certification of students. Unfortunately the practice has been to use the same test for multiple purposes, which is not always appropriate (Heubert & Hauser, 1999).

Assessment development. Psychometric theory and the process of test development form the basis of a quality assessment program (Moss, Pullin, Gee, Haertel, & Jones, 2008). The content and development process of current assessments are the by-product of theories of learning and measurement (NRC, 2001). Large-scale assessments developed to measure student achievement must move from the focus on discrete knowledge to more complex aspects of scientific literacy (Orpwood, 2001; Sadler & Zeidler, 2009). However, many of these methods

are not widely used because they are either not easily understood nor are they accessible to those without the technological infrastructure (NRC, 2001).

Dependable assessments come from test developers and curriculum developers working collaboratively with teachers (Darling-Hammond, 1994). Downing and Haladyna (2006) outlined *Twelve Steps for Effective Test Development* as a model for test developers to follow. The steps involve: developing an overall plan; defining the content and test specifications; developing the test items; designing, assembling, producing and administering the test; followed by scoring, testing for validity, reporting results. banking test items, and finally providing documentation in the form of a test technical report on the validity evidence for the test. Colorado convened content collaborative groups that followed a similar structure in their process of developing a resource bank of test items for Colorado districts to use as part of their comprehensive assessment program The content collaborative groups used an assessment review tool that set criteria for evaluating the alignment, scoring, fairness, and whether or not the assessment increases the student's opportunity to learn (CDE, 2014). The resource bank has been made available to districts, and the assessments within them are intended to serve as models for local assessment selections (CDE, 2014)..

Teachers have always been test developers and a component of their overall PCK includes their knowledge of assessment, involving their abilities to develop assessments that transform objectives into measurable assessment items. The ability to assess student gains towards learning goals is a crucial skill for any teacher, but Popham (1999) found that pre-service teachers graduate without an understanding or appreciation of assessment. Similar to the whole of PCK, which has been shown to develop and expand with years of experience, the results of a study by Mertler and Campbell (2005) found that in-service teachers scored higher

on an Assessment Literacy Inventory than pre-service teachers (in Siegel & Wissehr, 2011). With the exception of these few studies, very little has been published on in-service science teachers' knowledge and skills in developing common science assessments, especially those that are standards-based and district-wide.

Large-scale assessment systems. Large-scale assessment systems developed by state, national, and international testing companies have the benefit of trained psychometricians who test, edit and validate their items. However, most large-scale science assessments do not include items that measure extended thinking, reasoning skills, problem solving, or inquiry (Darling-Hammond & Pecheone, 2010; Quellmalz, Timms, Silberglitt, & Buckley, 2012). Herman, Heritage, and Goldschmidt (2011) claim that strengths and limitations of statistical models, which estimate the value individual teachers contribute to student achievement have been widely debated; yet, issues and practices concerning quality of the assessments have received little attention According to Downing and Haladyna (2006) a majority of the focus by the educational system and policy makers has been on the test scores and the data generated from these high-stakes achievement tests

Herman et al. (2011) argued that instrument validity is the overarching concept that determines whether an instrument can be considered *quality.* Five propositions outlined by Herman et al. (2011) formulate the evidence-based argument that the assessment measures what it is intended to measure and can be used to provide evidence for making decision about student achievment, including:

1. The standards clearly define what students are expected to learn.
2. The assessment instruments are *designed* to accurately and fairly address what students are expected to learn.

3. Student assessment scores accurately and fairly *measure* what students have learned.
4. Student assessment scores accurately and fairly *measure* student growth
5. Students' growth scores (based on the assessments) can be accurately and fairly attributed to the contributions of individual teachers (p. 3-4)

According to Herman et al. (2011), "Assessments in and of themselves are neither valid nor invalid. Rather, validation involves evaluating or justifying a specific interpretation(s) or use(s) of the scores" (p. 3).

Balanced large-scale assessment systems are needed to support high quality learning (Stiggins, 1999). Darling-Hammond and Pecheone (2010) cited the decline of the United States students on international assessment measures since the passing of NCLB in 2001. They call for development of assessment systems similar to high-achieving countries that require students to analyze, apply knowledge, and write extensively on primarily open-ended response items.

There has been an influx of technological advancements throughout our accountability systems. Quellmalz et al. (2012) developed and tested how simulation based assessments can become components of state assessment programs. The development involved six states and data analysis using Item Response Theory (IRT) showed high reliability and validity of the test items. IRT describes how well assessments work, specifically how well an individual item on an assessment works (Lord, 1980). Reliability measures indicate how consistently individuals respond to items, whereas validity indicates how accurately the item (or instrument) is measuring the construct of interest. Quellmalz et al. (2012) concluded that the simulation-based science assessments were successfully implemented across diverse settings, in addition, students scored higher on the computer simulation items and achievement gaps were lessoned, resulting in their

conclusion that a new generation of simulation-based science assessments have the potential to transform assessments systems

Assessment of scientific literacy. The Programme for International Science Assessment program (PISA) embodies the goals of science education reform, it represents an innovative approach to large-scale assessment systems that states and districts ought to try and emulate (Bybee, Fensham, & Laurie, 2009). PISA aligns their assessment with their vision for scientific literacy, which they define as the ability to use scientific knowledge and processes not only to understand the natural world, but also to participate in decisions that affect it. PISA is concerned with what citizens need to be able to do with science-related issues and they conclude that people most often have to make their own suitable conclusions from evidence. Citizens are faced with evaluating claims made by others on the bases of evidence, and they need to distinguish their personal opinions and values from evidence-based statements (OECD, 2009). While most standardized assessments use simple recall and application prompts, PISA aims to assess three competencies that are more process oriented: 1) Identifying scientific issues; 2) Explaining phenomena using scientifically correct language; and 3) Using scientific evidence (Sadler & Zeidler, 2009)

Despite the leadership of PISA and diligent research from the American Educational Research Association (AERA); The American Psychological Association (APA); and the National Council on Measurement in Education (NCME), which produced the Standards for Educational and Psychological Testing; high quality assessment systems aligned with high quality, rigorous standards still elude the education community. Orpwood (2001) argued, "leadership in assessment in support of curriculum change must come through research and the professional development of teachers, rather than through large scale international assessment

projects" (p. 136). While Steel, Hamilton, and Strecher (2010) in their report, *Incorporating Student Performance Measures into Teacher Evaluation Systems*, identified that locally developed assessments have the potential to be well aligned with local curricula, yet the items need to be developed, administered and scored in ways that promote high levels of consistency.

On a much smaller scale, Gormally et al. (2012) used an iterative process to develop the *Test of Scientific Literacy Skills* (TOSLS) in order to inform college biology instructors of their students' scientific literacy skills. The TOSLS is a freely available, 28-item, multiple-choice instrument. It was designed to specifically measure nine SL sub-constructs including evidence-based reasoning, examining the credibility of evidence, societal/technological implications, abilities and understandings of inquiry (including NOS), quantitative reasoning, transforming representations (graphing), statistical analysis and justification of data (Gormally et al. (2012). Instruments like the TOSLS show promise for assessment of scientific literacy, yet more research is needed to explore its' validity among science disciplines and populations beyond undergraduate biology courses.

Summary of the Problem

Educators, scientists, and policy makers are collaborating to help prepare teachers to help their students understand how scientific knowledge is generated and how to distinguish between scientific evidence and personal opinion (National Science Teachers Association [NSTA], 2010). Science educators in particular must possess an understanding how knowledge is generated, the nature of science, and what it means to be scientifically literate (NRC, 1996).

This study aims to identify whether there is an alignment between the scientific literacy skills identified in science education reform efforts (e.g., the standards teachers are expected to teach), teachers' perceptions of these skills, and teachers' abilities to demonstrate these skills

(through science assessments they develop for their students). The findings of this study will inform teacher educators and school administrators of the types of professional development opportunities teachers need. It is an opportune time to determine this alignment since the state of Colorado and many others are poised to begin evaluation of teachers based on student achievement. In lieu of state-developed assessments many content areas, including science, are developing district common assessments. These assessments are administered to students across the district all of which are registered for the same course, regardless of instructor or school.

Given the understanding that scientific and quantitative reasoning skills, specifically the use of evidence and reasoning in making claims and decisions are at the heart of scientific literacy; and the call for a focus on these skills is evident in national science education reform documents and standards (AAAS, 1990; Achieve, 2013a; College Board, 2009; NRC, 1996, 2007, 2012;); it would seem appropriate that the measurement of these skills would also be prevalent in our assessments that are measuring teacher effectiveness. The objectives of this study include uncovering the alignment between the skills called for in science education reform efforts and the assessments created by school district science teachers to measure student achievement.

CHAPTER 3: METHODS

The methodology involved in this non-experimental associative study begins with a restatement of the research questions, followed by a detailed explanation of the research design, including the multiple methods used to investigate each research question.

Research Questions

The overarching objective of this study was to determine the presence of alignment between what teachers know, what they are expected to teach, and how they assess scientific literacy. The following questions and sub-questions were posed in order to meet this research objective:

1. *How do science educators define scientific literacy?*
    a. How do Colorado science education reformers and policy makers define scientific literacy and how consistent is this with the definitions found in national science education reform documents?
2. *What are district secondary science teachers' perceptions of SL skills?*
    a. What skills do secondary science teachers perceive are necessary for their students to demonstrate scientific literacy? What are teachers' perceptions of when their students are SL? How do secondary science teachers rank the importance of aforementioned (Q1) SL skills?
    b. What is the alignment between teachers' ranking of importance of SL skills and their self-reported teaching/assessment of these skills?
3. *What is the relationship between teachers' overall levels of scientific literacy and the quality of common assessments they create?*
    a. How scientifically literate are district secondary science teachers?

b. What is the alignment between the SL skills in teacher assessments and SL skills embedded in CAS/NGSS?

c. Is the importance of SL skills, as reported by teachers reflected in the common assessments they create?

Research Design

A non-experimental associative research design was used to answer the research questions. An associative research approach was followed as the variables under study had many ordered levels (Gliner, Morgan, & Leech, 2009). Comparative and descriptive research questions were analyzed through inferential and descriptive statistics.

Variables

The content analysis, teacher survey, and TOSLS resulted in the following variables listed below according to research question:

RQ1: *How do science educators define scientific literacy?*

a. Frequency of SL skills in the State of Colorado Secondary Science Standards

b. Frequency of SL skills in the NGSSs.

RQ2: *What are district secondary science teachers' perceptions of SL skills?*

a. Science teacher perceptions of SL skills for their students

b. Teacher rank importance of SL skills

c. Teacher perceptions of when their students are scientifically literate.

d. Teacher frequency of Teaching SL skills

e. Teacher frequency of Assessing SL skills

RQ3: *What is the relationship between teachers' overall levels of scientific literacy and the quality of common assessments they create?*

a. Teacher level of SL as measured by the TOSLS

b. Frequency of SL skills in Common assessments

Context of Study and Sources of Data

This study was conducted in a single medium-sized school district. Poudre School District, situated in Northern Colorado, covers over 1800 square miles and includes the communities of Fort Collins, Windsor, Timnath, Wellington, LaPorte, and various mountain locations within Larimer County Colorado. With a total of three elementary school, 10 middle school, seven high school, two charter school, and one K-12 online school, the total sample population was 113 secondary science teachers (Poudre School District, 2010). The following sections highlight the data sources and samples that were used in the content analysis, teacher survey, and teacher *Test of Scientific literacy* or TOSLS. Content analysis was performed on two sets of secondary science standards: the State of Colorado Science Education Standards (CDE, 2009), and the Next Generation Science Standards (NGSS), (Achieve, 2013a), as well as six different grade-level teacher-developed district common assessments (Table 3).

In December of 2009, the Colorado Department of Education adopted pre-Kindergarten through 12th grade science standards. The developers of these standards embedded scientific inquiry and process skills into the three content standards (CDE, 2009). In addition, 21st century skills and post-secondary workforce readiness competencies were integrated into each grade level standard, including inquiry questions; relevance and application examples; and NOS (CDE, 2009). According to CDE (2009), the nature of the discipline highlighted in every subject area, includes "the characteristics and viewpoint one keeps as a result of mastering the grade level expectation" (p. 8). Though this definition does not correlate exactly with the scientific

communities' definition of NOS, the science standards do include multiple connections to scientific literacy skills within the NOS grade level expectations.

Table 3

List of Common Assessments Available for Content Analysis

Grade Level	Content Area	Time-frame
6th	3 tests: Physical, life, and earth systems science	End-of-unit
7th	3 tests: Life (cells, genetics) Earth (geology)	End-of-unit
8th	4 tests: Physical (energy & waves); Motion & forces; Earth (weather & climate); Astronomy	End-of-unit
9th	Biology	Year-end final
10th (primarily)	Earth systems science	Year-end final
11th (primarily)	Chemistry	Year-end final

In April of 2013, Achieve, Inc. unveiled the *Next Generation Science Standards*. The NGSS were developed through a collaborative effort across all 50 states. They were organized in a "coherent manner across disciplines and grades to provide all students an internationally benchmarked science education" (Achieve, 2013a, p. 1). The eight science and engineering practices, disciplinary core ideas, and cross-cutting concepts outlined in the National Research Council's *Framework* for K-12 Education provided the "blueprint" for the NGSS. The performance expectations in the NGSS's include the application of the science and engineering practices to content. Their intention is to focus on understanding and application rather the recall of facts (Achieve, 2013a). For this study, the NGSSs performance expectations will be analyzed for scientific literacy skills.

Poudre School District secondary science teachers who volunteered their time and expertise are the developers of the district common assessments under analysis. The assessments were designed to align with the 2009 Colorado Science Education Standards according to grade level and/or subject. Teachers used the evidence outcomes to drive their assessment development. The assessments have been used district-wide since the spring of 2011 at the end of the course (for high school) or the end of the unit (middle school) assessment as per their intention upon development. In the fall of 2012 science teachers were given the option to also pre-test their students on the common assessments so overall growth scores could be determined. Grade-level assessment teams have met during subsequent summers to analyze assessment data and make modifications to assessment items where applicable. The most recent version of each common assessment was analyzed for this study (Table 3).

Participants

Convenience sampling was used to recruit secondary science teachers to take a survey on skills and perceptions of scientific literacy. This non-probability sampling technique was appropriate since the participants were chosen on the basis of convenience and availability (Gliner et al., 2009). The population was drawn from both full- and part-time secondary science teachers contracted within Poudre School District during the 2013-2014 school year. With the support of district level personnel all 113 of the science teachers on record during the 2013-2014 school year were contacted via e-mail. Besides reducing the time and cost for data collection, the use of a web-based survey Qualtrics was chosen for this study because it can be designed for sequential responses (Evans & Mathur, 2005). In this study it was important that teachers answered the initial question on the survey regarding their definition of scientific literacy before proceeding with the survey during which they would read other definitions of scientific literacy.

One benefit of using an online survey was that, with the district administrative support, teachers were more likely to respond promptly and be forthright with their knowledge and perspectives.

Secondary science teachers were invited to participate in the research study through email. They received an introduction to the research project (including the option to consent to research, as required by both the university and school district institutional review boards), a link to the online survey (administered through Qualtrics), and information on how they could receive the teacher incentive offered to complete the survey. Potential participants were given two weeks to complete the survey, after seven days the district administrator (the science curriculum coordinator) sent out another e-mail invitation to remind teachers to complete the survey. At the two-week mark 21 surveys were completed. At that time, a third e-mail message was sent, this time by the Poudre School District Research and Development Coordinator who stressed the importance of the study for area teachers and district support staff in order to identify areas of potentially needed professional development. Another week was provided for teachers to complete the survey.

In order to increase the response rate an incentive to participate was included; this included a $20 coffee/travel mug value, paid for by the school district, which determined that the data collected from the survey would be valuable for their curriculum development and professional development efforts. A short focus-group interview with eight secondary science teachers at a district meeting early in the study revealed that teachers were more likely motivated to complete the survey if they knew that everyone would receive an incentive, as opposed to the chance of only a few individuals receiving an electronic incentive (e.g., Kindle, Fit Bit) if they entered their name in a raffle. At the end of the survey teachers were invited to provide their email addresses to receive the incentive. This project was not supported with extra-mural

funding, so we were limited in the incentives (provided at a discount by a local business) that we could offer teachers. Poudre School District personnel, including the district science curriculum facilitator and science department chairs from each middle school and comprehensive high school, supported the recruitment of teachers.

Methodology

Content analysis. Classical content analysis (Nuendorf, 2002) guided the examination of the standards and common assessments. Content analysis may be briefly defined as the systematic, replicable technique for compressing many words of text into fewer content categories based on explicit rules of coding (Krippendorff, 1980; Nuendorf, 2002). Content analysis can be either quantitative (also known as manifest content analysis) or qualitative (latent content analysis) (Schreir, 2012). These two categories have also been defined as conceptual analysis and relational analysis. Conceptual content analysis establishes the frequency of concepts found in the text (quantitative), whereas relational analysis examines the meaning among concepts in a text (qualitative) (Busch et al., 1994-2012). Krippendorf (2004) suggested that the "quantitative/qualitative distinction is a mistaken dichotomy between the two kinds of justifications of content analysis designs… For the analysis of texts both are indispensable" (p. 87). Schreir (2012) cited that this "sharp contrast becomes blurred on closer inspection" (p. 14). A qualitative approach to content analysis requires taking context into account, while at the same time attempting to balance both quantitative and qualitative research (Schreir, 2012). For the purpose of this study, Nuendorf's (2002) definition and method of content analysis was used:

> Content analysis is a summarizing, quantitative analysis of messages that relies on the scientific method (including attention to objectivity, inter-subjectivity, *a priori* design, reliability, validity, generalizability, replicability, and hypothesis testing) and is not

limited as to the types of variables that may be measured or the context in which the messages are created or presented (p. 10).

A quantitative content analysis using an *a priori* coding framework was completed to determine the frequency of scientific and quantitative reasoning skills implicit in grade 6-12 Colorado Academic Standards in Science (released in December 2009) and the Next Generation Science Standards (released in 2013). This analysis was completed to highlight the importance of the inclusion of these skills in both science curricula and assessments. In addition, a similar content analysis of teacher developed district common assessments was conducted in order to determine the amount of focus scientific literacy skills have been given in district-level *high-stakes* assessments.

Constant comparative analysis. To assess teachers' perceptions of science literacy, an open-response survey was administered. The data collected from the survey allowed me to determine science teachers' definitions of scientific literacy, their self-reported frequency of both teaching and assessing various scientific literacy skills, as well as how they ranked the importance of the same skills. These data were compiled and coded using the inductive process of constant comparative analysis (Strauss & Corbin, 1990) for the open and axial coding levels. The selective or final codes were determined using the *a priori* coding framework of the content analysis once it was verified that teachers had identified all the skills on the framework. This allowed me to better answer the alignment questions of this study.

To assess teachers' scientific literacy levels, the *Test of Scientific Literacy Skills* (TOSLS) (Gormally et al., 2012) was inserted after the SL teacher survey. The exact questions including graphics were entered into Qualtrics online research suite survey software with permission from the TOSLS authors (Gormally, personal communication). Qualtrics software

was convenient since it allowed for the use of various question types, imports of multiple file types, and data exporting. Furthermore, these data could easily be shared with the school district, which was helping recruit participants, supplied incentives, and had a vested interest in the findings.

Instrumentation

The multiple method design of this research study included the use of various instruments. These included a coding framework to conduct the content analyses of academic standards and district common science assessments; a teacher perception survey to measure participants' views of scientific literacy and their teaching and assessment of SL skills; and the TOSLS, an instrument designed by Gormally et al. (2012) to measure participants' abilities to apply their SL skills. The instruments and their respective validity and reliability measures will be presented.

***A priori* coding framework development**. The content analysis *a priori* coding framework was developed specifically for this study after a thorough review of the literature was completed for chapter two of this document. The framework was cross-referenced with the nine scientific literacy skills outlined by Gormally et al., (2012). Weber's (1990) procedure for developing an *a priori* coding framework begins by establishing categories based on theory. In this manner categories are "a group of words with similar meaning or connotation" (Weber, 1990, p. 37). Professional colleagues should agree to the categories before the coding is applied to the data. Revisions and tightening of the categories should be mutually exclusive and exhaustive (Stemler, 2001; Weber, 1990). Developing the categories for the scientific reasoning portion of the coding frame drew on Toulmin's (1958) *Patterns of argumentation* framework, *The Claim, Evidence, and Reasoning Framework* (CER) by McNeill & Krajcik (2012), and the

Evidence-based Reasoning Framework (EBR) by Brown et al. (2010). Category refinement and final definitions were supported with the *Critical Thinking* and *Quantitative Reasoning* VALUE rubrics that were created and published by The Association of American Colleges & Universities (AAC&U, 2012).

To minimize the categories within the coding framework scientific explanation and argument were conflated into a single code, whereas within the literature one can find separate definition even though the delineation between the two is difficult. For example, Sampson and Clark (2009) describe an explanation as a statement that describes natural phenomenon. Furthermore, McNeill and Krajcik (2012) describe it as the inclusion of a claim, evidence, and reasoning. A scientific argument, on the other hand according to Driver et al. (2000) is a process of knowledge construction in which individuals clarify, critique, construct, and revise ideas in an effort to make sense of the natural world, and according to Osborne and Patterson (2011) an argument is a product that seeks to justify an explanation or to persuade someone.

The categories of the QR portion of the coding frame were based on contemporary literature, specifically: the PISA assessment framework (OECD, 2009), *Taking Science to School'*s four strands of science proficiency (NRC, 2007), and the three dimensions of the 2012 Framework for K-12 Science Education (NRC, 2012). All of these reports focus on Science, Technology, Engineering, and Math (STEM) practices, an important component of which is QR (Duschl, 2012). Additionally, the AAC&U (2012) *Quantitative Literacy* VALUE rubric specific to quantitative literacy was consulted since their definition of Quantitative Literacy (also known as QR) guided this study.

Initial drafts of the coding frame were then compared to the nine scientific literacy skills highlighted in the TOSLS instrument. Eight out of the nine skills TOSLS developers saw as

important were already included in the initial *a priori* coding framework. It was determined by the researcher and inter-rater coders to include Gormally et al.'s (2012) ninth skill (performing and interpreting statistical analysis of data). Basic statistical computations (mean, median and mode) were added to an already existing code "mathematical computations," while "understanding and interpreting basic statistics" was added as an additional code. However, after all coding was complete there were so few codes under *Interpret basic statistics* that all scores in that category were combined with code *Mathematical processes*.

The initial coding framework developed for this study included 14 separate skills, five more than the TOSLS mainly due to the separation of the codes "Claim," "Evidence," and "Reasoning;" and the addition of a code related to the use of models or reference to modeling. The final coding framework adjusted after correlations with TOSLS included nine total categories (Tables 4 & 5).

Coding framework validity. Efforts were made to ensure data collection with the coding rubric had high construct validity, meaning the rubric truly measured the theoretical constructs of scientific and quantitative literacy. The relevant literature and the nine skills tested on the TOSLS provided the background necessary for the development of the coding framework that was used in this study. To increase the validity of this instrument, each draft of the *a priori* coding rubric was critiqued from two university PhD level science educators and two secondary science teachers each having least twenty years of experience. Modifications were made until there was 100% agreement among the experts that the coding framework was all encompassing of scientific literacy skills.

Table 4

Evolution of coding framework

Initial Coding Framework (14 codes)	Adjustment after Correlations with TOSLS	Notes
1. Claim 2. Evidence 3. Reasoning	1. Claim	Still coded for claim if assertion does not include reasoning. Evidence fits with other codes depending on context.
1. Scientific Argument / Explanation	2. Scientific Argument / Explanation	
3. Evaluate the validity of sources	3. Evaluate the validity of sources	
4. Societal; Technological implications	4. Societal; Technological implications	
5. Methods of Inquiry	5. Methods of Inquiry	
6. Transforming representations	6. Transforming representations	This code combined two TOSLS codes into one (Interpreting graphs and creating graphical representations)
7. Mathematical processes	7. Mathematical processes (Including Interpret basic statistics)	This code also combines Interpreting basic statistics with Mathematical processes.
8. Analyze 9. Interpret 10. Evaluate	8. Justify conclusions based on quantitative evidence	Depending on the item initial codes in this category fit our definition of justification. Codes that involved interpretation of data were coded as (5/6 or 7/8).
11. Models or Modeling	9. Models or Modeling	This code was not a part of TOSLS, but an important component of SL according to research
10. Interpret basic statistics	Combined with Mathematical processes	

Table 5

Final a priori *Coding Framework Used for Content Analysis*

Code	Definition	Coding Rules
1. *Claim*	A. An assertion or answer to a scientific question. Communicating science knowledge.	1) Includes conclude, predict, hypothesis, infer.
2. *Scientific Argument / Explanation*	B. Includes claim, evidence, and reasoning; identify or evaluate the validity of an argument; recognize when evidence supports a hypothesis	2) Explain; explanation
3. *Evaluate the validity of sources*	C. Takes into account source of information based on scientific criteria, e.g. methods, peer-review, funding source	3) Conduct literature review; examine bias in media
4. *Societal / Technological implications*	D. Relating to science, technology, and society; socio-scientific issues; appropriate use of science to make societal decisions	4) Reference to greater community/technology
5. *Methods of inquiry*	E. Ask scientific questions; carry out scientific investigations; identify strengths and weaknesses of research design related to bias, sample size, experimental controls	5) Inquiry question; inquiry lab
6. *Transforming Representations*	F. Create graphical representations from data table and vice versa	6) Create/read/interpret graphs
7. *Quantitative Skills*	G. Basic mathematical operations; use of formulas; solve problems using quantitative skills (includes interpreting basic statistics)	7) Includes a concept that involves mathematical calculations
8. *Justification*	H. Justify inferences, predictions, conclusions based on quantitative data	8) Beyond just analyzing data, look for claims based on data.
9. *Models or Modeling*	I. Abstract representation of phenomena or data; can be theoretical or mathematical in nature	9) Reference to/analysis of existing model; create a model

Coding framework reliability. Efforts were made to ensure that this data collection instrument – the *a priori* coding framework had a high level of reliability. Reliability of the instrument was ensured through multiple *pilot* studies pre-testing the coding framework. Originally the framework consisted of 14 categories separated into scientific and quantitative reasoning. Two raters/science educators were trained on the use of the coding framework and each of them completed content analysis of 20% of pre-selected standards and common assessment items approximately 15 from each data set. Initial coding rounds brought about many questions and even additional categories for the rubric. Subsequently, the raters were trained on the final version of the coding framework, including clarification of the definitions, examples, and all acceptable synonyms. Raters were provided with their own excel spreadsheet to enter their frequency counts independently. Reliability was calculated using a Cohen's Kappa, which takes into account the fact that raters are expected to agree with each other a certain percentage of the time simply based on chance (Cohen, 1960 as stated in Stemler, 2001. Initial (pre-) Cohen's Kappa reliability measures ranged from 50% in the high school NGSSs to 75% in the middle school NGSS's. Following discussion post- Cohen's Kappa measures ranged from 84% again in the high school NGSS to 100% in the NGSS middle school Standards (Table 6).

Table 6

Cohen's Kappa Values for Inter-rater Reliability of Standards

Standards	Pre (%)	Post (%)
Colorado High School	72.8	97.6
Colorado Middle School	54.1	94.3
NGSS High School	50.0	84.1
NGSS Middle School	74.9	100.0
Average	63.0	94.0

Teacher Survey

The teacher survey used throughout this research study was adapted from the test of science literacy (TOSLS) with only minor modifications to content and format. Gormally et al. (2012) used multiple means to establish content validity when developing the TOSLS, including a faculty survey about skills essential for scientific literacy. The TOSLS faculty survey was used to "verify the consistency of the skills (they had) articulated through their literature review" (Gormally, et al., 2012, p. 366). The process completed by the authors of the TOSLS was very similar to what I used for validation of the *a priori* coding rubric developed for the content analysis of this research study (e.g., literature review and consultation with expert faculty members).

Specific modifications to the TOSLS university faculty/professor survey included the addition of a key question that asked teachers to identify how they know when their students are scientifically literate and the three most important scientific literacy skills their students should master. Moreover, yes/no prompts within the original survey were changed to Likert scale responses. Within the series of questions in the original survey regarding each of the scientific literacy skills tested, Gormally et al. (2012), asked yes/no questions of their participants (on whether they currently taught/assessed a certain skill in their classes). For my study, I chose to change the yes/no prompts to Likert scale, which allows each individual to express him or herself on the extent to which they teach, assess, and rank the priority of a particular scientific literacy skill (McLeod, 2008).

The modified TOSLS survey provided the needed data for the variables being tested. These include teacher: demographics (level and years taught); perceptions of a scientifically literate student and skills students should master for scientific literacy; and self-reported

frequency of teaching, assessing, and ranking of importance of scientific literacy skills. Questions related to teachers' experience and abilities related to common assessment development were also added to the original TOSLS faculty survey. Two questions from Gormally, et al.'s (2012) original teacher survey that related to respondents ranking of the importance and incorporation of skill vs. content learning in the classroom were removed from the teacher survey in this study since the questions were not related to variables being tested. See Appendix A for the complete TOSLS teacher survey used in this study.

Teacher survey validity. Since the instrument used in this research strand was slightly modified from its original form, and the intended audience was different from the original study, it was recognized that instrument validity was threatened. The procedure that was followed to improve the internal validity included a thorough peer review of the teacher survey prior to its final publication. Researchers and district personnel who were not part of the participants in this study, but who possessed the similar demographics to study participants were chosen to provide feedback on the instrument itself. Initial readings both verbal and silent of the survey were completed and peer feedback was received and recorded from all participants offering comments. Minor adjustments and all edits were made to the survey until all feedback was addressed and peer reviewers were satisfied with the product.

Teacher survey reliability. Since the teacher survey was administered only once to teachers, neither test-retest nor parallel forms reliability was performed. Comparing the teacher responses to similar items in Gormally et al.'s (2012) faculty survey supported the internal consistency reliability of the survey.

Test of Scientific Literacy Skills (TOSLS)

Gormally et al. (2012) designed an instrument for college science faculty members to measure their students' knowledge and skills of scientific literacy. Item difficulty analysis and item discrimination indices on the completed TOSLS indicated "it is meaningful to view a student's score on the TOSLS as a measure of his or her scientific literacy skills" (Gormally et al., 2012, p. 373). Because the instrument was developed for use with adults who were in the process of completing their undergraduate degrees in the life sciences, it was deemed appropriate for the use with science teachers, who must have earned at least a bachelor's degree in science to work as a public teacher in Colorado. The TOSLS is freely available to the public, and it is in a format that is easy to administer and score. The TOSLS contains 28 questions correlated to nine scientific literacy skills that were pre-determined by a thorough literature review and verified by expert agreement (Appendix B).

TOSLS validity. The developers of TOSLS undertook a thorough iterative process, using multiple means to determine instrument validity, and focusing mainly on measures of content and construct validity (Gormally et al., 2012). However, the TOSLS was originally developed for undergraduate biology students, thus the instrument validity was checked since it was now being used with in-service science teachers. Classical test theory (Lord, 1980) was used to determine whether the instrument was functioning the way it should with the new population. Two tests were completed with the teacher data: *Percent Correct* for each item and a *Discrimination* test. If there was a very high level or a very low level of percent correct the item needed to be considered if it was providing sufficient information, a discrimination test was then performed that the item was of appropriate difficulty. In the discrimination test, teacher data were scored, entered into a data spreadsheet, and responses were ordered from high to low.

Examining each question, the percent correct for the bottom third of respondents was determined and subsequently was subtracted from the percent correct from the top third of respondents. A discrimination value between 0.2-0.5 was considered acceptable (Lord, 1980).

Data Collection

Data collection consisted of content analyses of state and national standards and district common assessments using the *a priori* coding framework; frequency statistics of teacher demographics from the survey; constant comparative analysis (Strauss & Corbin, 1990) of open response survey items; and scoring of the TOSLS using the scoring guide provided by Gormally et al. (2012).

Content analysis. The most current copies of state (December 2009) and national standards—Next Generation Science Standards (Achieve, 2013a) and district common assessments were used for this study. Preparing the standards for analysis involved removing extraneous statements while maintaining the structure provided. Within the state standards the data were removed from the state document and organized in a spreadsheet. The numbering system of the standards was maintained both for grade level and content standard, while the evidence outcome (EO) and NOS statements were transferred onto the spreadsheet (Figure 5). The number of statements that were coded was 153 for High School State Standards (EO's n=99; NOS statements n=54) and 164 for Middle School State Science Standards (EO's n=101; NOS statements n=63).

The Next Generation Science Standards have been organized into student performance expectations with supporting *foundation boxes,* which include the science and engineering practices, disciplinary core ideas, and cross cutting concepts (Achieve, 2013a; Figure 6). Within the middle school and high school standards each performance expectation was copied into an

Content Area: Science
Standard: 1. Physical Science

Prepared Graduates:

- Apply an understanding of atomic and molecular structure to explain the properties of matter, and predict outcomes of chemical and nuclear reactions

Grade Level Expectation: High School

Concepts and skills students master:

4. Atoms bond in different ways to form molecules and compounds that have definite properties

Evidence Outcomes	21st Century Skills and Readiness Competencies
Students can: a. Develop, communicate, and justify an evidence-based scientific explanation supporting the current models of chemical bonding b. Gather, analyze, and interpret data on chemical and physical properties of different compounds such as density, melting point, boiling point, pH, and conductivity c. Use characteristic physical and chemical properties to develop predictions and supporting claims about compounds' classification as ionic, polar or covalent d. Describe the role electrons play in atomic bonding e. Predict the type of bonding that will occur among elements based on their position in the periodic table	**Inquiry Questions:** 1. How can various substances be classified as ionic or covalent compounds? 2. What role do electrons play in different types of chemical bonds? **Relevance and Application:** 1. Related compounds share some properties that help focus chemists when looking for a substance with particular properties for a specific application. For example, finding new super conductors. 2. Carbon atoms bond in ways that provide the foundation for a wide range of applications. For example, forming chains and rings such as sugars and fats that are essential to life and developing synthetic fibers and oils. 3. Living systems create and use various chemical compounds such as plants making sugars from photosynthesis and chemicals that can be used as medicine, and endocrine glands producing hormones. **Nature of Science:** 1. Recognize that the current understanding of molecular structure related to the physical and chemical properties of matter has developed over time and become more sophisticated as new technologies have led to new evidence. 2. Employ data-collection technology to gather, view, analyze, and interpret data about chemical and physical properties of different compounds.

Figure 5. Example: Colorado Academic Standard (CDE, 2009)

2.Structure and Properties of Matter

2.Structure and Properties of Matter

Students who demonstrate understanding can:

2-PS1-2. Analyze data obtained from testing different materials to determine which materials have the properties that are best suited for an intended purpose.* [Clarification Statement: Examples of properties could include, strength, flexibility, hardness, texture, and absorbency.] [Assessment Boundary: Assessment of quantitative measurements is limited to length.]

The performance expectations above were developed using the following elements from the NRC document *A Framework for K-12 Science Education*:

Science and Engineering Practices	Disciplinary Core Ideas	Crosscutting Concepts
Analyzing and Interpreting Data Analyzing data in K–2 builds on prior experiences and progresses to collecting, recording, and sharing observations. ▪ Analyze data from tests of an object or tool to determine if it works as intended. (2-PS1-2)	**PS1.A: Structure and Properties of Matter** ▪ Different properties are suited to different purposes. (2-PS1-2)	**Cause and Effect** ▪ Simple tests can be designed to gather evidence to support or refute student ideas about causes. (2-PS1-2) ***Connections to Engineering, Technology, and Applications of Science*** **Influence of Engineering, Technology, and Science on Society and the Natural World** ▪ Every human-made product is designed by applying some knowledge of the natural world and is built using materials derived from the natural world. (2-PS1-2)

Connections to other DCIs in second grade: N/A

Articulation of DCIs across grade-levels: **5.PS1.A** (2-PS1-2)

Common Core State Standards Connections:

ELA/Literacy –

RI.2.8 Describe how reasons support specific points the author makes in a text. *(2-PS1-2)*

W.2.7 Participate in shared research and writing projects (e.g., read a number of books on a single topic to produce a report; record science observations). (2-PS1-2)

W.2.8 Recall information from experiences or gather information from provided sources to answer a question. (2-PS1-2)

Mathematics –

MP.2 Reason abstractly and quantitatively. *(2-PS1-2)*

MP.4 Model with mathematics. *(2-PS1-2)*

MP.5 Use appropriate tools strategically. *(2-PS1-2)*

2.MD.D.10 Draw a picture graph and a bar graph (with single-unit scale) to represent a data set with up to four categories. Solve simple put-together, take-apart, and compare problems using information presented in a bar graph. *(2-PS1-2)*

Figure 6. Example: Next Generation Science Standard organization (Achieve, 2013a).

excel spreadsheet for analysis. A total of 130 performance expectations were coded from the NGSSs (71 High school and 59 Middle school).

After permission was granted from the Poudre School District, teacher-developed common assessments from six science courses, including middle school science (grades 6, 7, and 8), and high school biology, chemistry and earth systems science, which are used by teachers across the district, were analyzed. These common assessments were prepared for analysis by removing the multiple-choice selections leaving just the stem of each question to be coded. Where applicable instructions on the use of the images/graphs/readings was maintained and provided for inter-rater coders where applicable. The same *a priori* coding framework was used to analyze the assessment items, along with the procedure for determining inter-rater reliability (e.g. separate analysis followed by discussion and resolution of differences).

Teacher survey. All 113 secondary science teachers in the district were invited to take the online teacher survey, which generated 49 completed surveys through the Qualtrics survey program. The data collected were exported into a data management spreadsheet (Excel or SPSS) for further analyses. Qualitative coding of teacher perceptions of scientific literacy using constant comparative analysis was completed. Inter-rater reliability was completed on 20% of the open-response items with one-university faculty and two-master level secondary science teachers (one middle school and one high school), neither of who was involved in this study nor were recruited to complete the survey. The inter-rater coding process resulted in Cohen's Kappa of 96.5%, which can be considered almost perfect agreement (Landis & Koch, 1977).

TOSLS. The final piece of data collection came from the administration of the TOSLS to the science teachers. Following the teacher survey items on Qualtrics, the teachers were

introduced to the TOSLS and the 28-item instrument was presented to them in the exact same format provided by the developers.

Data Analysis

Descriptive statistics were completed on the content analysis to determine the frequency of scientific literacy skills in both the district developed common assessments and the aforementioned standards. Initially the standards and assessment documents were coded as either a yes (1) or no (0) under the applicable category(ies) in the coding framework. Some standard statements or test items were very specific and fell within a single code. For example, "Describe for various waves the amplitude, frequency, wavelength, and speed" was coded as *Claim.* Other items were more complex and had multiple "yes" codes (1s). For example, "Select and use technology tools to gather, view, analyze, and report results for scientific investigations about the characteristics and properties of waves" was coded as *Scientific Arguments/Explanations*, *Societal/Technological Implications;* and *Methods of Inquiry*. Since each item must yield a single score point the multiple scores were equally distributed across the cells in the content analysis. For example, initially a standard statement may have received "yes" codes in three categories (*Claim*, *Transforming Representations*, and *Mathematical Processes*), the score for this statement was equally distributed across the three cells identified (Table 7). According to Porter, Smithson, Blank, and Zeidner (2007), the process of re-distribution and equally weighting multiple codes is a major assumption, which may not reflect what the authors of the standards or assessments intended.

Table 7

Sample Coding and Equal Weighting of Codes for Standards

Standard	Claim	Transforming Representations	Mathematical processes (including stats)	Total
1.1 Gather, analyze and interpret data and create graphs regarding position, velocity and acceleration of moving objects	(1) 0.333	(1) 0.333	(1) 0.333	(3) 1.0

A quantitative measure of the alignment between proportions of: state and national standards; standards and teacher perceptions; and standards and teacher developed common assessments was calculated using the Porter Alignment Index (Porter, 2002). The alignment index was calculated as follows:

$$\text{Alignment Index} = 1.0 - \frac{\Sigma|x-y|}{2}$$

where x and y are cell proportions in the two different data sets. Calculation of the alignment index can result in values from 0 (no alignment) to 1.0 (perfect alignment) and the higher the value the better the alignment.

Porter's alignment index has been used as a quantitative measure of the alignment between proportions of: State and national standards, and standardized tests (Liu & Fulmer, 2008); State by State comparison of Standards (Porter et al., 2007); State Standards and State Exams (Contino, 2013); and Intended, Planned, and Enacted general and special education curriculum (Kurz, Elliot, Wehby, & Smithson, 2010). Porter et al. (2007) checked alternative indices and concluded that they are all "nearly perfectly correlated" with the Porter Index (p. 35). To alleviate any confusion between the Porter Alignment Index (P) and p-values determined for statistical significance (p), the Porter Alignment Index will hereafter be referred to as the PI

through the remainder of this document. One application of the PI is described by Liu et al. (2009), they used the Porter Alignment Index to determine the alignment between physics content standards and nationally standardized tests. Liu et al. created an algorithm to determine alignments between pairs of tables and with 20,000 alignment measures determined that "an alignment of 0.780 is needed in order to be statistically significant at the .05 level" (Liu et al., 2009, p. 781).

Teacher survey. The data generated from the survey was analyzed using two strategies: a qualitative analysis of open-ended questions using constant comparative analysis and a descriptive and associative statistical analysis of survey items using Likert scales. A constant comparative analysis was used to develop selective codes of each teacher's perception scientific literacy. Participants' responses were also analyzed to determine their respective definitions of scientific literacy. In addition, participants were asked to consider whether they teach, assess, and value the same set of scientific literacy skills that were counted in the standards and assessments content analysis. The following variables generated from the survey were analyzed: a) teachers' SL perceptions; b) teacher demographics; c) teachers' self-reported frequency of teaching SLSs; d) teachers' self-reported frequency of assessing SLSs; e) Teachers' ranks importance of SLSs; and f) teachers' experiences with district common assessment development.

Upon completion of the qualitative coding of teacher perceptions of scientific literacy using constant comparative analysis, a content analysis of the responses was completed using the same *a priori* coding framework of the standards. This framework considered acceptable since teachers had already identified all the SL sub-constructs I had coded for in the standards. In doing this, I was able to compare groups of teachers using various demographic data with their

self-reported frequency of assessing scientific literacy skills. I inductively looked for patterns in the descriptive statistics that could be reported.

TOSLS. Scoring of teacher's level of scientific literacy using TOSLS was completed using the scoring guide provided by the instrument developers (Gormally et al., 2012). Overall level of scientific literacy of teachers in the district of study was determined.

Establishing Trustworthiness

There are four ways of establishing trustworthiness according to Guba (1981):

1) Credibility (in preference to internal validity)

2) Transferability (in preference to external validity/generalizability)

3) Dependability (in preference to reliability), and

4) Confirm-ability (in preference to objectivity).

This chapter highlighted how internal and external validity was ensured with expert agreement on SL skills in the *a priori* coding framework; training standards to inter-rater coding; peer review of survey contents; and the use of classical test theory on the TOSLs results. Reliability or consistency measures included: pilot studies of the *a priori* coding framework, and comparison of teacher survey responses to university faculty members' responses. Efforts to maintain objectivity were supported by school district administration of the survey and communication with participants.

Limitations

The major limitations of this study were the small sample size and the narrow demographic region selected for this study. This will affect my ability to generalize and apply results to other districts in the state and nationally. Another limitation of this study is the survey results were based solely on participating teachers' self- reported frequencies of both teaching

and assessing scientific literacy skills; this may have influenced results of research question two (determining the difference between teachers ranking of the frequency of importance versus assessment of scientific literacy skills). Despite the localized data collection, the study findings will be of interest to educators and administrators around the state and nation. The dilemma of how to assess scientific literacy skills across districts and to assess teachers on their performance teaching scientific literacy is timely and relevant. There are few studies that have addressed this issue in part because the NGSS were only released in March 2013.

The use of Porter's Alignment Index in this study also had limitations given that in the studies cited (Contino, 2013; Liu & Fulmer, 2008; Liu et al. 2009; Porter et al., 2007) there were five cognitive levels the researchers coded for in their content analyses, whereas in my study I coded for nine science literacy sub-constructs. As a result of having more coding units, the overall Porter's alignment indices, which are based on the sum of all the differences, could be skewed towards being less aligned.

CHAPTER 4: FINDINGS

The overarching objective of this study was to determine the alignment between what teachers perceive and know; what they are expected to teach; and how they assess scientific literacy. To meet this objective, it was imperative to determine whether science teachers 1) demonstrate scientific literacy knowledge and skills; 2) self-report to teach and assess scientific literacy skills and, 3) have developed district assessments that are aligned with scientific literacy reform goals and documents.

Of the 113 secondary science teachers in one Northern Colorado school district who received the invitation to participate in this study, 43% completed the survey portion (n=49), 25% of the teachers completed the survey, the TOSLS, and the questions related to district common assessment development (n=28). Out of the possible 67 women and 46 men who received the survey 43% of each gender responded. An average of 43% of district employees in three of the four "Years of teaching experience" categories responded. The one category that had no participants was teachers with 4-5 years of experience. While the district employs 9 teachers in that category 0% of the 9 participated in the survey. Education level was more variable with a 29% response rate from teachers with a Bachelor's degree, 51% of the district teachers with a Master's degree participated and the one person in the district with a PhD participated making that category 100% participation. Lastly, 51% of district Middle School and 35% of High School Science teachers participated in the survey. (Table 8)

Table 8

Demographics of Survey Participants

Demographic Category	Number of Survey Participants	% of Total District Science Teachers (n=113)
1. Gender	Male – 20 Female – 29	Male – 43% Female – 43%
2. Years Teaching Experience:	0-3 years – 7 4-5 years – 0 6-9 years – 10 10+ years - 32	44% 0% 40% 44%
3. Education Level:	BS/BA = 13 MS/MA = 35 PhD/EdD = 1	29% 51% 100%
4. Primary Teaching Level:	Middle School = 26 High School = 22	51% 35%

Data Analysis

RQ1: How do science educators define scientific literacy? How do Colorado science education reformers and policy makers define scientific literacy, and how consistent is this with the Next Generation Science Standards?

The purpose of this analysis was to determine if the Colorado Academic Standards (CASs) are in line with the NGSSs. A Porter's alignment index (PI) was calculated to determine the extent of alignment. Using Liu et al.'s (2008) critical significance value of 0.78, all sub-categories in the coding frameworks for both High school (Table 9) and Middle School (Table10) were aligned. The overall Porter's alignment index between the CAS and NGSS High school was 0.778, which when rounded off is equal to the significance level of 0.78. However,

the CAS and NGSS Middle School (PI = .688) cannot be considered aligned based on Liu et al.'s (2008) significance level of 0.78.

Table 9

Alignment of CAS and NGSS for High school

SL Sub-constructs	CAS HS Frequency/n	NGSS HS Frequency/n	Absolute value of difference	PI
Claim	.163	.243	.080	.960
Scientific Argument / Explanation	.280	.262	.018	.991
Source Credibility	.099	.038	.061	.970
Society/Tech.	.127	.058	.069	.966
Methods of Inquiry	.095	.046	.049	.976
Transforming Representations	.039	.011	.029	.986
Mathematical Reasoning (including basic statistics)	.086	.126	.040	.980
Justification (Numerical)	.056	.093	.037	.982
Models or Modeling	.053	.115	.062	.969
Overall Porter's alignment index			Σ = .444	PI= 1-(Σ Differences /2) = .778

Table 10

Alignment of CAS and NGSS for Middle School

SL Sub-constructs	CAS MS Frequency	NGSS MS Frequency	Absolute Value of Difference	PI
Claim	.299	.132	.167	.917
Scientific Argument / Explanations	.252	.377	.125	.937
Source Credibility	.084	.021	.063	.969
Society/Tech. implications	.121	.058	.063	.969
Methods of Inquiry	.0762	.062	.014	.993
Transforming Representations	.0091	.010	.001	1.00
Mathematical Reasoning	.0467	.083	.036	.982
Justification (Numerical)	.0212	.143	.111	.939
Models or Modeling	.0832	.116	.033	.984
Overall Porter's alignment index			Σ= .624	PI = 1-(Σ Differences /2) =0.688

The CAS at the high school level is better aligned to NGSS HS than the CAS MS is aligned to the NGSS, specifically for the sub-constructs (1) *Scientific Arguments / Explanations* for which CAS had a PI of .991, whereas for the MS the PI between the CAS and NGSS was .937. Both CAS and NGSS had the greatest difference between them in the sub-construct, *Claim,* with CAS having higher frequencies then the NGSS. In addition there were some middle school sub-constructs that could be considered "higher-ordered thinking skills," including, *Scientific*

argument / Explanations and *Justification*, within these areas, the CAS had lower frequencies than NGSS.

The survey contained a question on whether or not teachers had read the NGSSs. Thirteen teachers (26%) answered "yes," 16 teachers (32%) identified that they had "scanned their contents" and 21 out of the 50 teachers (42%) responded, "I have not read them." Teacher's who responded with a "yes" or "have scanned their contents" were asked to respond to this question: "In your own words, what are the primary features that differentiate the NGSS's from previous standards (e.g., National Science Education Standards, Colorado Science Education Standards)?" Teacher responses were analyzed using constant comparative analysis (Table 11).

RQ2a: What are district secondary science teachers' perceptions of SL skills?

To answer this question several analyses were conducted. The iterative process of constant comparative analysis (Glaser & Strauss, 1967) was initially used to analyze the open-ended responses in the teacher SL survey in order to determine secondary science teachers'1) perceptions of necessary skills to demonstrate scientific literacy and 2) perceptions of assessing scientific literacy. The ranking of both aforementioned scientific literacy skills, including teacher frequency of teaching and assessing these skills, was compiled and analyzed with descriptive frequency statistics.

Initial coding of teacher responses to two survey items: Question 7) How do you know when your students are scientifically literate?, and Question 8) What are the three most important scientific literacy skills for students to master?, resulted in 40 open codes. Tallies of demographics were concurrently recorded. Through inductive coding (Strauss & Corbin, 1990), thirteen axial codes were identified for question 7, and ten axial codes were identified for question 8 after collapsing codes. At this point, it became apparent that science teacher

respondents' perceptions overlapped with the codes in the coding framework developed for analysis of the state and national standards. With the exception of two additional codes for survey questions 7 ("related to test scores" and "unsure") and 8 ("critical thinking" and "other"). These findings are of interest because teachers who participated in this study identified all of the SL skills that are identified in the standards and science education reform literature. It is also interesting due to the additional categories that emerged.

Table 11

Selective Codes for Teacher Perceptions of how the NGSSs Differ from Previous Standards

Selective Codes	Example Teacher Responses
Practices rather than content/facts	"The NGSS are performance based instead of knowledge based."
Science knowledge supported by evidence	"(They are) designed to show how scientist develop a knowledge base that is supported by evidence."
Increase in inquiry	"No longer *the scientific method*"
Connection to CCSS in Math and English Language Arts	"The science content is not different, but the addition of including literacy in the form of reading and writing and the inclusion specifically of technical writing is important…"
Connection/application to engineering and technology	"I believe that the connections made specifically to technology and engineering are the difference."
Connection to practical uses; broader community	"The NGSS is intended to show the interconnections of scientific content, application in the scientific community and preparation for students to succeed at advanced levels of education or careers."
Grade level specific expectations	"Grade level specific expectations."
Change in number of standards and GLEs	"Change from 5 standards to 3; Number of grade level expectations went from 155 to 82."

Subsequently, all science teacher responses to both survey questions (7 & 8) were recoded using the *a priori* content standard coding framework (Table 5). By doing so, the language of analyses of teachers' perceptions of scientific literacy skills was consistent with the analysis of both the content standards and teacher-developed common assessments. Moreover, using the *a priori* coding framework developed for this study helped establish it's external validity and hence, the trustworthiness of the findings (Table 12).

Table 12.
Selective and Axial Codes from Constant Comparative Analysis of Survey Questions 7 & 8.

Axial Code	Selective Code = SL sub-construct	Sample Teacher Narratives to Survey Q8
Claim, evidence reasoning	*Scientific Arguments*	Write a scientific conclusion with a claim, supporting evidence and reasoning.
Content knowledge; Communicate	*Articulating Claim*	Rich vocabulary (to help with reading/internet search)
Source Credibility	*Evaluation of the Validity of Sources*	Evaluation of scientific *findings* for credibility
Application – real world; NOS	*Societal / Technological Implications* (Includes NOS)	Application / apply to real-life
Inquiry	*Methods of Inquiry*	Valid experiments / Controls
Data Transformations	*Transformation of Representations*	Representing and interpreting graphical data
Quantitative Reasoning	*Mathematical Reasoning (including basic statistics); Justification; Models or the Process of Modeling*	Manipulate equations; Connect concrete data and abstract ideas; Interpreting and creating models
Additional axial and selective codes for survey Q7		
Related to Test Scores	Content mastery / test proficiency	Assessment success of 70 / 80% or higher.
Unsure	Unsure	What is scientifically literate?
Additional axial and selective codes for survey Q8		
Critical Thinking	Coded as combination of Scientific Argument and Methods of Inquiry per Glaser (1941)	Thinking critically when learning or performing science
Other	Unsure	Depends on the course

Though the responses contained skills consistent with science education reform documents, it is important to note the frequency range of responses. For example, only one teacher answered "interpreting and creating models" models in response to this item, and eight responses (the majority of them MS science teachers) included "writing scientific conclusions with claim, evidence, and reasoning," which could be coded as *scientific argument* or *explanation*. The greatest frequency of responses (.79) were categorized as, "Communicate with appropriate science content language," which could be categorized as *Claim* in the coding framework. These responses were equally divided between MS and HS teachers and ranged from lower cognitive responses such as science vocabulary and identify the main idea from text to what could be considered a higher cognitive demand task such as communicating understanding in various ways (Bloom, 1967).

Responses to the prompt "How do you know when your students are scientifically literate?" were initially open coded using constant comparative analysis. It became apparent during the coding process that the axial codes reflected the various levels of Bloom's taxonomy of cognition (Bloom, 1967) and that the selective codes again overlapped with the coding framework developed during the review of content standards. Once again, the responses were recoded using the initial coding framework to determine the alignment of teacher responses to the categories identified in the standards. Similar to the SL skills prompt, the most frequent teacher response was related to *Science Content Knowledge* and communicating that knowledge (Table 13). Axial codes relating to *content knowledge* included: *Basic knowledge* (identifying vocabulary in text and describing scientific phenomena), *Comprehension level* responses

(explaining concepts, answering 'leveled' questions, speaking with appropriate language), and *Analysis level* responses (discuss relationships among key terms and deciphering the main idea of text)

Table 13

Axial and Selective Code Results from Survey Question 7 "How do you know when your students are scientifically literate?"

Axial Code	SL sub-construct	Frequency/ Category		Sample teacher narratives (SQ7)
		MS	HS	
Communicate evidence/results	Scientific arguments/ explanations	.19	.15	Make evidence-based conclusions/communicate w/evidence
Knowledge Comprehension Application	Articulating a claim	.88	.50	Answer leveled questions; decipher the main idea
Source credibility	Evaluate the validity of sources	.12	.05	Evaluate/question validity of (material)
Application	Societal/ technological implications (includes NOS)	.19	.15	Apply knowledge to everyday life; knowledge can change as new evidence is uncovered
Inquiry/finding answers Scientific investigations	Methods of inquiry	.35	.05	Ability to research/experiment in order to determine the answer
Interpret/infer	Transforming representations	.15	.23	Interpret data graphs, tables, diagrams
	Justification	.15	.27	Explain relationship between variables
	Models or modeling	.04	.05	Interpret/create models
Analysis of quantitative data	Mathematical reasoning (includes basic statistics)	.19	.09	Analyze data (material)
Unsure	Unsure	.15	.23	
Related to test scores	Content mastery/ test proficiency	.12	.32	

After teachers answered the open response items related to their perceptions of SL, they were asked to identify the frequency that they teach and assess the SL sub-constructs identified in this study. Subsequently, they ranked the importance of the SL sub-constructs from not important at all to the most important skill overall. The responses included "low level importance" of three skills (*Evaluating the validity of sources*, *Methods of inquiry,* and *Modeling*) and one respondent identified skill three (*Societal/ technological implications*) as "not important." Teachers ranked the skills *Reading and interpreting graphical data*, *Justification based on quantitative data*, and *Creating the appropriate graphs from data*, the highest respectively (combining "very important" and "the most essential skill overall" choices) (Figure 7).

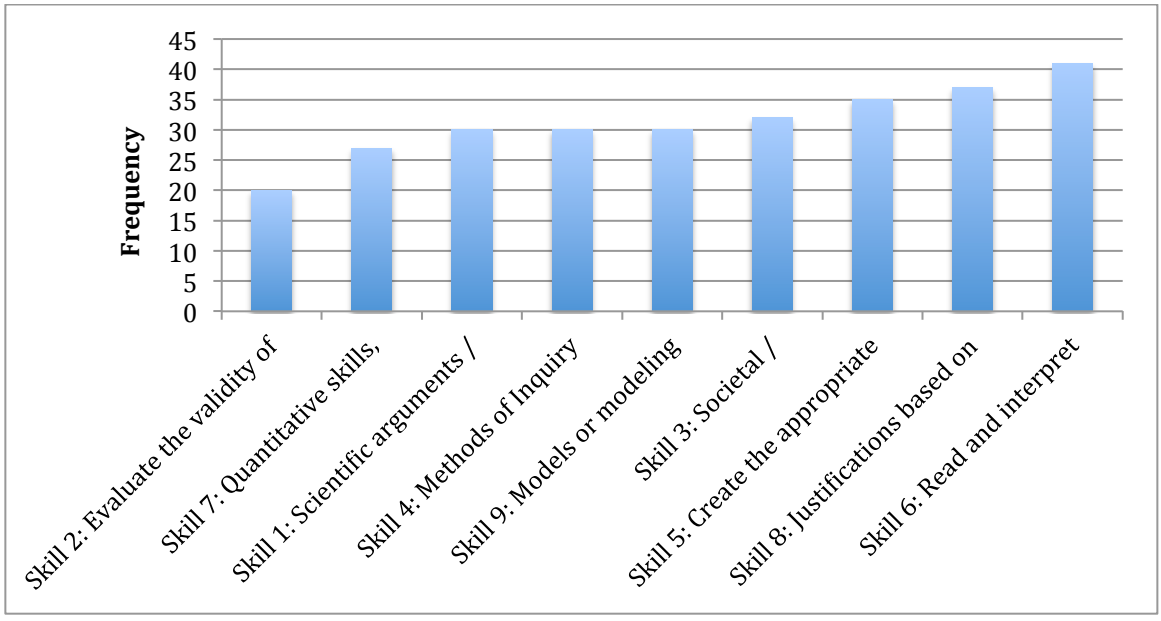


Figure 7. Frequency of teacher ranking of the importance of the SL Skills. Frequency found from combining "very important" and "the most essential skill overall" choices.

RQ2b. What is the alignment between teachers' ranking of importance of SL skills and their self-reported teaching/assessment of these skills?

Ordinal variables with five ordered levels require the use of a Spearman Rank Order Correlation Coefficient or Spearman Rho. There were no significant relationships between any pair of variables ($p < .050$). The teacher ranking between importance and whether or not they assess the skill of Societal and Technological implications ($r= -.90$; $p = .08$) had a large negative correlation coefficient, which would indicate that teachers ranked the importance of this skill high, while the assessment of this skill was ranked towards the lower end of how frequent it was assessed.

Overall modes and means for teacher ranking of importance of each skill was above the ranking of teaching and assessing every nine skills; however, none of the value differences were run for significance given the Spearman correlations results run previously. (See Appendix F for complete analysis of RQ2c)

RQ3. What is the relationship between teachers' overall levels of scientific literacy and the quality of common assessments they create?

RQ 3a. How scientifically literate are district secondary science teachers?

The average TOSLS scores out of the 28 teachers that completed the survey was 84.9%, with a range from 64.3% to 100%. Middle School and High school teachers were within 0.1 % of the average. While the male to female scores were 88.1to 82.5% respectively and master's level versus bachelor's degree scores were 85.6 to 82.1%. When compared with test subjects by Gormally et al., (2012), teachers fell below the average of college Biology professor's 91% average, and above all sub-category of undergraduate Biology major's 66% average, thus it can be concluded that teachers are scientifically literate.

The TOSL instrument was used to assess each SL sub-construct on multiple items, and the average percent correct for each sub-construct (SL skill) was calculated. The two lowest skills included *Methods of Inquiry* and *Create the appropriate graph from data* (Table 14).

Table 14

Teacher Proficiency Scores on Sub-constructs of SL

	TOSLS item # (% correct)	Average % correct
Skill 1: Scientific arguments / Explanations	1(100), 8(85), 11(81)	89
Skill 2: Evaluate the validity of sources	10(81), 12(88), 17(83), 22(100), 26(74)	85
Skill 3: Societal / Technological Implications	5(97), 9(96), 27(95)	96
Skill 4: Methods of Inquiry	4(84), 13(85), 14(38)	69
Skill 5: Create the appropriate graph from data	15(65)	65
Skill 6: Read and interpret graphical representations	2(74), 6(100), 7(87), 18(83)	86
Skill 7: Quantitative skills, including basic statistics	16(91), 20(88), 23(96)	92
Skill 8: Justifications based on quantitative data	3(90), 19(83), 24(70)	81
Skill 9: Models or modeling	21(100), 25(83), 28(77)	87

Percent correct analysis of the TOSLS test revealed four items that all the teachers got correct. These four items were spread across four sub-constructs: identify a valid scientific argument; evaluate the validity of sources; read and interpret graphical representations of data; and justify inferences based on quantitative data. On the other hand, one item in particular related to identifying strengths and weakness of experimental design had the greatest variation in responses. Question 14, which was the last in a series of questions related to the strength of research design related to a study on health effects of drinking diet soda, asked: "Which of the following attributes is **not** a strength of the study's research design? This question fell under Skill 4: Methods of Inquiry, and only 38% of the responses were correct (Figure 8).

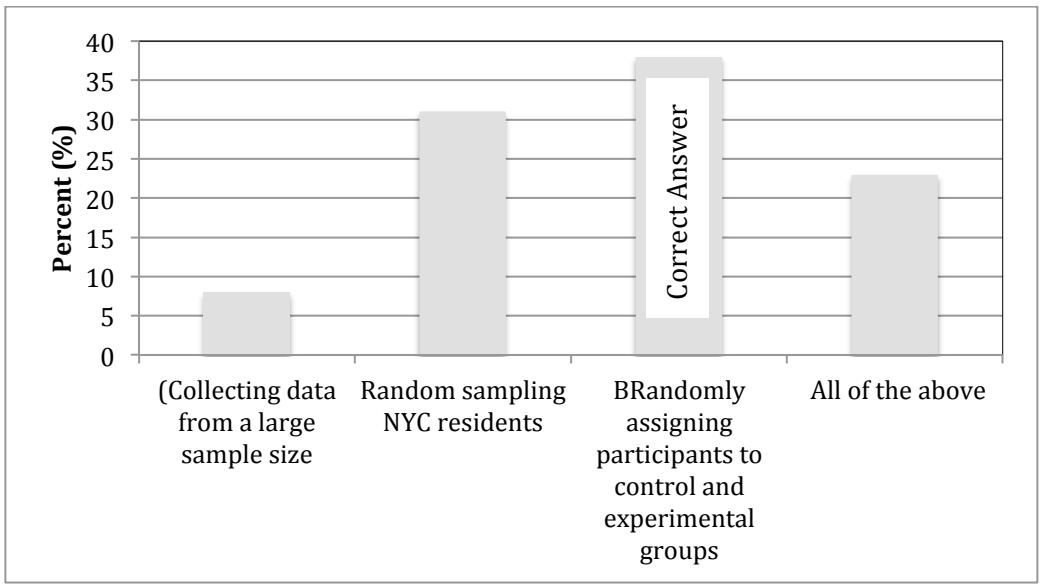


Figure 8. Frequency of various responses from TOSLS test question number 14

A discrimination test was performed on the TOSLS scores. 16 items did not fall within the range of 0.2-0.5, which is considered acceptable (Lord, 1980). After removing the 16 items and re-calculating the scores, there was a decrease in the overall teacher level of scientific literacy by 3% down to 81.8%. Cronbach's alpha levels were calculated to determine the internal consistency reliability of the TOSLS. Overall the entire test had an alpha of .93, which is considered excellent (Gliem & Gliem, 2003). Cronbach's alpha for individual SL sub-constructs ranged from unacceptable levels for skills 3 & 6 (α= .413 & .491) to good for skill number 7 (alpha = .820).

RQ3b. What is the alignment between the SL skills in teacher assessments and SL skills embedded in the specific Colorado Academic Standards (CAS) they were written to address?

Common Assessments were coded in the same fashion as the CAS and NGSSs using the *a priori* coding framework. The code frequencies for each grade level unit or course common assessment were compared to the CAS they were written to address. Porter's alignment index

values range from 0.29 for both the 6th grade Earth Science exam and the 9th grade Biology exam, to 0.51 for the 7th grade combined exams on Cells and Genetics (Table 15), which means they are not aligned at the 0.78 level

The least aligned sub-category was 1. *Claim*. Every grade level assessment except seventh grade was not aligned (at the 0.78 significance level) in that subcategory, with the teacher developed assessments having higher frequencies of *Claim* questions than the standards they were designed to address. Seventh grade assessments sub-categories aligned in all areas including *Claim* with a PI = .80 for the Cells and Genetics tests and PI = .87 for the Earth Science Geology end of unit exam. See appendix G for the grade level assessment data for all SL sub-constructs aligned with CAS.

Table 15

Common Assessment Alignment Indices

Grade Level	Content Area	Porter's Alignment Index (PI)
6th	Physical science	.32
	Life science (ecology)	.38
	Earth science	.29
7th	Life science (cells and genetics)	.51
	Earth science (geology)	.50
	Physical science (motion & forces, and energy	.47
8th	Earth science (weather and climate)	.42
	Earth science (astronomy)	0.34
9th	Biology	0.29
10th	Earth systems science	0.32
10th-11th	Chemistry	0.39
Average PI for all assessments = 0.38		

Four open-ended questions were presented at the end of the TOSLS to teachers that identified that they participated in the district common assessment development process. Open and axial coding was completed on teacher responses to questions directed at them personally such as: "Describe your experience and knowledge of developing common science assessments;" "Describe any professional development experiences you have had with instrument or assessment development;" and "Why did you choose to participate in the PSD common assessment development process?" Selective codes mimicked the axial codes for the majority of these questions (Table 16) this is possibly due to the small sample size (n=16).

Table 16

Teacher Experience Related to Common Assessment Development

Survey Question	Axial/Selective Code	Teacher Narrative
Describe your experience and knowledge of developing district common science assessments.	- Exams are content focused - Emotional - Increase awareness of district-wide focus - Lack of plan for test improvement and validity	- Limited to multiple choice - Tedious and frustrating - Aware of other schools / teachers' focus
Describe any professional development experiences you have had with instrument or assessment development.	- None - Lessons on Depth of Knowledge - During formal education training - Through (the years) collaboration with peers	- None - District facilitator took us through 2 hour lesson - I had two lessons on how to write a test back in 1987 - I help create every common assessment in my courses.
Why did you choose to participate in the PSD common assessment development process?	- Professional responsibility - Concern of how tests will be used for teacher evaluations - Instructional Planning - Voice / input on exam	-Critical professional interaction - My effectiveness as a teacher will be evaluated using this tool - It helps drive my instruction - Important that I had a say in what was being tested.

The direct question: "Describe the process of common assessment development," was analyzed by identifying events in a sequence rather than a traditional constant comparative method. A sequential description of how district educators developed their common course assessments and teacher narratives are outlined below (Table 17).

Table 17

Sequential Process of District Common Assessment Development

General Outline	Teacher Narrative
Standards were divided up among teachers that gathered to participate for each common course.	– We looked at standards – Depending on the group size individual teachers or teacher partners wrote or designed questions that aligned to the Evidence Outcomes in the Standards. – We look at the Colorado standards and wrote assessment questions that would show mastery of the standard. – Developed questions that progressively move toward advanced categories for the standard being assessed.
Shared questions with the whole group	– Through consensus determined the critical knowledge and process in which the students needed to be assessed.
Made modifications	– But not many – When editing, we tried to assess the level of the skill that the question was asking, and tried to find a level I, II, or III (Depth of knowledge) per area of focus on the state standards. – Developed rubrics that delineate levels of proficiency
Made sure all the standards were covered in the tests	
Piloted some of the tests (2012)	
Came back the next summer to look at the test again and made some changes.	– Some of the tests were piloted and have been revised. Others are still in the first year of development – Revised questions every year

Summary

The evidence from this study identified three major findings (1) CAS can be considered aligned to the NGSS at the high school level, but the middle school CAS are not aligned to the

middle school NGSS; (2) Teacher perceptions of SL are diverse in nature and contained all the SL skills identified in the reformer's literature. However, the frequencies at which some of the skills were identified did not align with reformer's perceptions; and (3) Teacher developed common assessments were not aligned to the standards they were intended to measure. They tended to contain much more lower cognitive level questions related to science content.

CHAPTER 5: DISCUSSION

With the publication of Science Education National Reform documents such as *A Framework for K-12 Science Education* (NRC, 2012) and the *NGSS* (Achieve, 2013a) science teachers in some states are explicitly addressing the science and engineering practices, as well as disciplinary core ideas and crosscutting relationships in their respective classrooms. Despite efforts of national science education reformers and initial positive responses from local teachers, the shift from school and district-level accountability policies to state-level teacher evaluation policies are likely to impact teacher behavior and ultimately the curricular and instructional choices science teachers make. As student performance on standardized assessments plays a more central role in the teacher evaluation process, building a test worth teaching to is important (Briggs, 2013; Wagner, 2008).

The objective of this research study was to determine the alignment between what teachers perceive and know about scientific literacy (SL); what SL skills they are expected to teach; and how they assess SL (Figure 9). As teachers feel compelled to respond to administrators and policy makers who argue for both pedagogical and content changes, they are also concerned about the impact that student performance may have on their own ratings as effective (Anderson, 2012). Hence, although teachers may have a positive attitude regarding curricular change, their behaviors may not reflect changes in practice because of these extenuating circumstances, as social cognitive theorists might note (Bandura, 1977). The results of this study can guide researchers in a better understanding of teacher perceptions of reform movements in order to better support science teachers (teacher educators; professional development providers; administrators, etc.). To do so, though, researchers and administrators must first recognize the conflicting policy constraints that teachers are under.

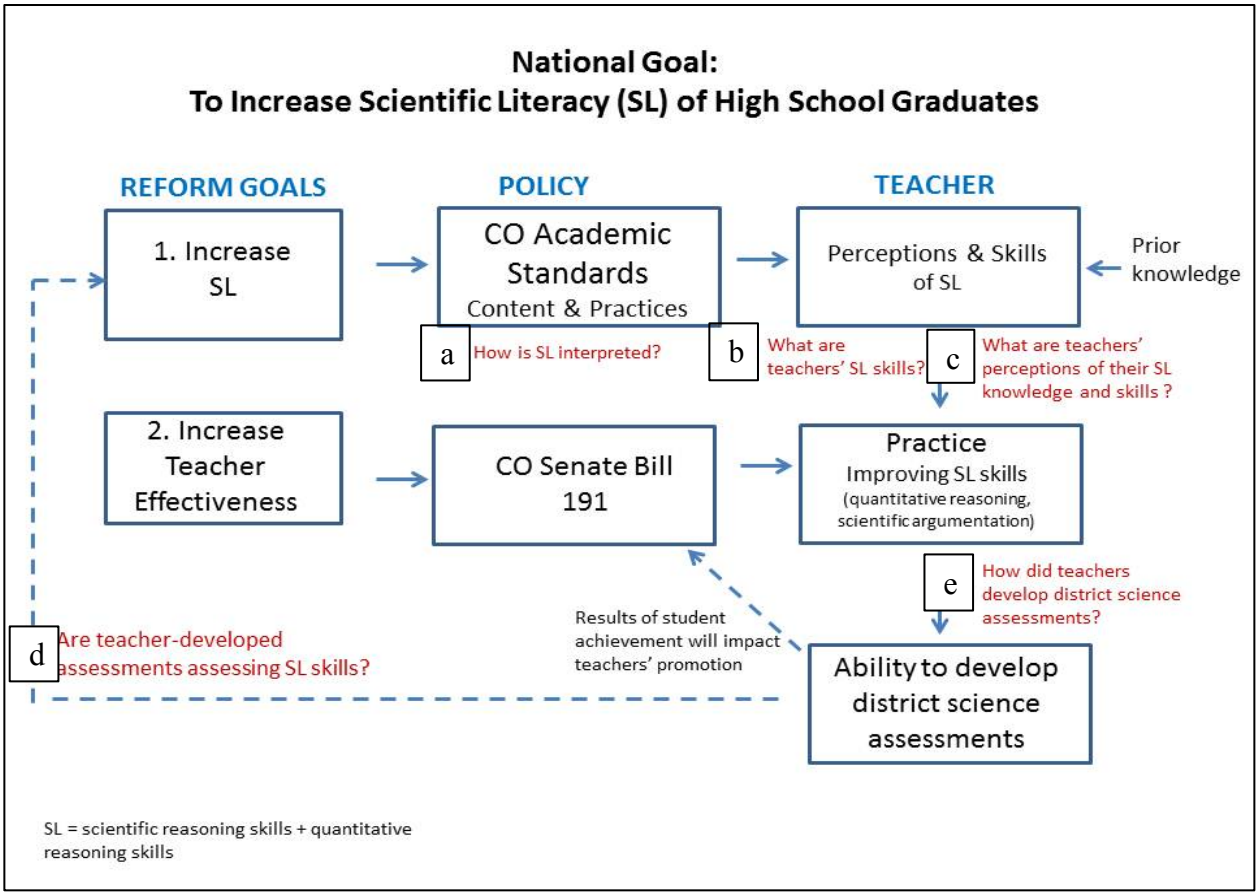


Figure 9. *Conceptual Framework highlighting key findings of this study*

My intentions in this research study were to answer the following questions:

1. How do science education reformers define SL?
2. What are district secondary science teachers' perceptions of SL skills?
3. What is the relationship between teachers' overall levels of SL and the quality of common science assessments they create?

The findings demonstrated that the Colorado Academic Standards for SL skills are aligned with the national Next Generation Science Standards (NGSS) at the HS level, and almost in alignment with NGSS at the MS level (Figure 9a). In addition, teacher participants demonstrated sufficient SL skills (Figure 9b). However, teacher perceptions of SL skills were not aligned with reformer's perceptions included in both the CAS and the NGSS (Figure 9c). The teacher developed

common assessments were not in line with SL reform documents (Figure 9d), and teachers involved in the common assessment development process expressed frustration, lack of professional development, and knowledge that the tests were content-based and not assessing SL skills sufficiently (Figure 9e).

Alignment between Standards, Teacher Perceptions, and Practices

This study described the alignment among reform goals, teacher perceptions of reform, and teacher skills to apply reform goals. This research approach is informative because a central tenet of science education reform is there should be alignment between content standards and high-stakes standardized tests (Liu et al., 2008). Concurrently, it is known that teacher knowledge, skills, and abilities underlie their behavior and what is ultimately implemented in the classroom (Larkin, Seyforth, & Lasky, 2009; Woodbury & Gess-Newsome, 2002). National and Colorado state level science standards are not in complete alignment, which poses a problem for policy makers and school administrators who expect teachers to adhere to national science education reform goals. This study has demonstrated that national reform documents, specifically the NGSSs (Achieve, 2013a), place a greater emphasis on the skills related to scientific arguments /explanations than the Colorado Academic Standards (CAS). In addition, CAS at the middle school level are not aligned to the middle school level NGSS (Tables 9 and 10). Currently, science teachers in Colorado are expected to teach and are being evaluated on state standards; therefore, there are few incentives for Colorado teachers to meet national standards. The current study's findings are consistent with Liu and Fulmer's (2008) analysis of the alignment between the science curriculum and assessments in NY State Regents Exams. They found that while there was high alignment between the content of curriculum (state standards) and the Regents test, there were considerable differences of cognitive demand

between the two. Woodbury & Gess-Newsome (2002) argued that teacher practice is likely to change only in response to changes in attitudes that are, in turn, shaped by a hierarchical level of contextual factors (departmental, school, district, state, and national). Because Colorado's policy for evaluating teacher effectiveness impacts teachers' job security, it is likely that local and regional factors will trump national factors.

Although the Colorado Department of Education has announced that it will not adopt the NGSS until 2018 at the earliest, Colorado educators, including the participants in this study, will likely feel pressure to be a) prepared for reform efforts coming in the near future (the neighboring district, Thompson School District, is already piloting the NGSS) and b) in alignment with other reform efforts that are advocated by the National Science Teachers Association and the Association of Science Teacher Education.

Science teacher perceptions of SL encompass the skills outlined in science education reform definitions but at varying frequencies (Tables 14 and 15). Teachers in this study perceived SL as the attainment of a rich content vocabulary that would allow their students the ability to read and write scientifically and technically. They were less likely to describe SL as the ability to argue with evidence using scientific and quantitative reasoning. In addition, approximately 20% of teacher perceptions uncovered that they were unsure of what SL meant for themselves or their students. In other words, the definition of SL was not something they felt they could articulate. Several teachers expressed that student performance on standardized or classroom assessments was an indicator of SL, a definition not shared by science education reformers. The dilemma, therefore, is whether standardized tests do in fact assess students' abilities to demonstrate understanding of how science knowledge is generated and disseminated, and the critical skills related to inquiry, scientific argumentation, quantitative reasoning.

Kang, Orgill, and Crippens (2008) concluded that there is "a gap between the teachers' conceptions of inquiry and the ideals for the reform movement" (p. 337), as I have also concluded. Using multiple measures (individual response and group responses to inquiry scenarios) Kang et al. (2008) studied 34 high school science teachers' conceptions of scientific inquiry and found that their perceptions rarely mentioned evaluating and communicating explanations (claims, evidence, and reasoning). The teachers in my study did mention abilities in communicating scientific arguments and explanations as an indicator of SL; but they did so at a much lower frequency than skills that required a higher cognitive demand.

Though the teachers in my study described knowledge and comprehension of science content as indicators of SL, they also ranked the *reformer skills* as either "very important," or "the essential SL skill over all the others" when surveyed. This is consistent with Gormally et al.'s (2102) identification of the disconnect between college level instructors' value of increasing their students' SL, their teaching of SL skills, and the assessment of their students' SL skill proficiency. College professors' open-ended responses regarding the most important skills for SL mirrored the responses from teacher participants in this study. In hindsight, if the teachers would have been asked to rank SL skills in order from most to least important, I could have identified the skills deemed least important by them.

Breaking down teacher perceptions by demographic variables indicated that HS teacher perceptions of the most important SL skills were those that are concrete and quantitative (graphing and interpreting data). MS teacher perceptions, on the other hand, were more abstract and included specific references to the CER framework regarding how to write scientific conclusions and justification of results (Table 13). The fact that MS teachers were also more likely than their HS counterparts to perceive a student's abilities in inquiry as an indicator of

their SL level is important. There is no consistent use of the term, inquiry, which has made it difficult to measure teachers' perceptions of it. Therefore, the NRC (2012) recommended the use of the term, scientific practices, to replace scientific inquiry, in hopes of better aligning teacher perceptions with those of science education reformers. Teachers' perceptions of SL are likely influenced by their understanding or lack of understanding of NOS. Roehrig and Luft (2007) found that novice high school teachers did not demonstrate understanding of SL because of their limited knowledge of NOS. Windschitl (2003) found that knowledge of NOS and likelihood of using inquiry-based instructional strategies could be explained by authentic research experiences that teachers had while in college. Hence, if reformers hope for teachers to increase their own SL, it is important for them to use consistent language and to assist in teachers' development of a strong understanding of NOS.

Teacher participants in this study were assessed as having higher average TOSLS scores than the undergraduate biology majors in Gormally et al.'s (2012) study. Although the teachers participants were less scientifically literate (average score of 85%) compared to college biology professors (average score of 91%), this is not surprising considering the majority of science teachers in the study have not had science research experiences at the graduate level and only one of the teacher participants had doctoral degrees. Teacher scores on the TOSLs identify that the majority of teachers that completed the test have sufficient SL skills, but there is room for improvement and development of these skills. Gormally and colleagues developed the TOSLS as an instrument to measure expected skills for undergraduate science students. It is reasonable to expect science teachers to be able to demonstrate the same knowledge and skills.

Teacher developed common assessments primarily contained content level questions, which were present at a higher frequency than the state academic standards they were written for.

There were also fewer questions requiring reasoning, and justification of responses. Similar studies by Contino (2013) examining the alignment between NY state core Earth Science curriculum and the state Regents exam found a Porter's Alignment Index to be 0.35, which was determined to be being slightly aligned based on Liu and Fulmer's (2008) methodologies. Contino (2013) also found that the Regents Exam had many questions coded at the *Remember* cognitive level, unlike the state core curriculum with which they were supposed to be aligned. In contrast, Liu and Fullmer's (2008) examination and alignment indices between the NY core Physical Science curriculum and Regents tests were high with an index around 0.8 for Physics and 0.7 for chemistry. Both of these studies analyzed cognitive skill levels of Performance indicators compared to the Regents Exams designed to assess them. It is unclear why the alignment indices are so variable between the different content areas of the two studies.

What is noteworthy, though, is that researchers in New York have begun to address the important issue of standardized state assessment alignments with standards. Because each state is able to interpret the national content standards, teachers from around the country are held accountable to slightly different content standards. Furthermore, if standardized assessments are not consistently addressing the same SL sub-constructs, then the purpose of having national academic standards may be mute. Indeed, the language used in the previous national science education standards (NRC, 1996) has been criticized for being too authoritative and hence, misleading about how science discourse occurs (Wallace, 2012). Furthermore, Wallace (2012) argued, standards may be written in a way that actually poses a barrier for classroom teachers who are expected to teach about how science knowledge is generated when they cannot replicate this in their own classrooms.

Defining Scientific Literacy (SL)

SL, within science education research, can be categorized based on different sub-constructs: (1) Fundamental Language Arts Skills (reading and writing; Norris & Phillips (2003)); (2) Science Content knowledge, which is common to all definitions of SL (Roberts, 2007); (3) NOS and knowledge generation through science inquiry (AAAS, 1990; Alchin, 2011); (4) Science evidence-based decision-making (National Science Education Standards, NRC, 1996); and (5) Scientific Reasoning (including quantitative reasoning, arguments, and evidence-based reasoning) (NRC, 2007; 2012). Often when researchers study SL they do not always acknowledge the variations in the definitions used by others. Science teacher organizations, such as the National Science Teachers Association (NSTA), do not always distinguish between these various definitions. In fact, because science education researchers also publish in this practitioners' journal, several SL definitions are represented in their high school magazine—*The Science Teacher* (See Metz, 2006), and their middle school magazine—*Science Scope* (see LoGiudici & Ende, 2010). It is, therefore, essential to examine teachers' perceptions of what SL is and to then evaluate how they apply this definition in their own assessment tools for their respective students.

In this study the evaluation of teacher SL was conducted through the triangulation of three data sources: (1) teacher perceptions of SL; (2) teacher knowledge or competencies of SL; and teacher skills or application of SL. Analysis of the intersection between SL constructs and teacher perceptions, knowledge, and skills, can shed light on future theoretical studies of how research can inform teaching and assessment practices (Table 18).

Table 18

Intersection of SL Constructs and Teacher Perceptions, Knowledge and Skills

(✓indicates presence according to my findings. + indicates frequent presence, – indicates low frequency; ± indicates present, but at a range of cognitive levels*)*

SL Constructs	**Teacher Perceptions**	**Teacher Knowledge**	**Teacher Skills** (As demonstrated by common assessment development)
Fundamental Language Arts Skills	✓±	? Unknown abilities	(✓) Assessments contained minimal technical reading with comprehension
Science Content knowledge	✓+	(✓+) Teacher performance on TOSLS indicated presence of content knowledge and SL skills of argumentation, NOS, and decision-making	(✓+)Assessments more frequently measured content knowledge
Scientific Reasoning Argumentation	✓		(✓−)Assessments were less likely to measure SL skills of argumentation, NOS, and decision making
NOS	✓−		
Decision-making	✓−		

Teachers' perceptions, knowledge, and skills around Fundamental Language Arts Skills are noteworthy. Teacher survey data indicated a range of cognitive levels (from knowledge to application level) related to reading and writing in science (hence, the ✓± rating; see chapter 4 Tables 13 and 14). What follows is virtually an unknown understanding of teacher knowledge and skills of the fundamental language of SL reform. More data are needed to capture more explicitly what teachers believe fundamental literacy means as it relates to SL. Moreover, teachers' own knowledge and skills around fundamental literacy (syntax, grammar, vocabulary) needs to be explored further if any conclusions are to be drawn. Halliday and Martin (1993) argued that people use language to exchange meaning, so it is necessary to determine exactly

what teachers mean when they express their ideas about SL. In my study I found that teachers exhibited regular use of SL reform language, but they were unable to apply that understanding when they developed district content-specific SL assessments. This begs the question of whether teachers are all interpreting the SL reform documents in the same way. One might speculate if they have developed their own nuanced definition of the terms because there are few opportunities for teachers to explore the meaning of reform language.

Studies that examine how people (in this case, reformers and teachers) use language when writing or speaking about science are important if reformers hope for gains to be made in science education reform. Wallace (2004) proposed a theoretical framework for researching literacy and language use in the science classroom (Figure 10). The framework identifies three sub-constructs: (1) Authentic science communication, which can range from the use of a personal vernacular to the correct usage of scientific vocabulary in expression; (2) Multiple discourses, which range from first hand private observations to the use of only authoritative public explanations in communication; and (3) Bhabha's "Third Space" (1994), which describes how speaker/writer and listener/reader co-construct meaning. This framework is useful in the analysis of teachers' knowledge, perceptions, and skills related to SL sub-constructs. In other words, when examining the overall teacher responses, I was able to determine where along the three continua—*voice, expression,* and *meaning making*—teachers lie. Although teachers *expressed* themselves using "scientific" language (i.e., reform language, such as "argumentation" or "scientific reasoning") and public *voice* (i.e., imperative, authoritative statements about what SL means to them without describing limitations of their ideas), their ideas about SL were not clear about how they make meaning of this construct. I conclude this because there was a misalignment between their perceptions and knowledge, and their abilities to apply their

knowledge when constructing assessment tools. I argue that this can be, in part, explained because teachers have limited opportunities to explore the multiple meanings of SL within their peer groups and with science education researchers and reformers.

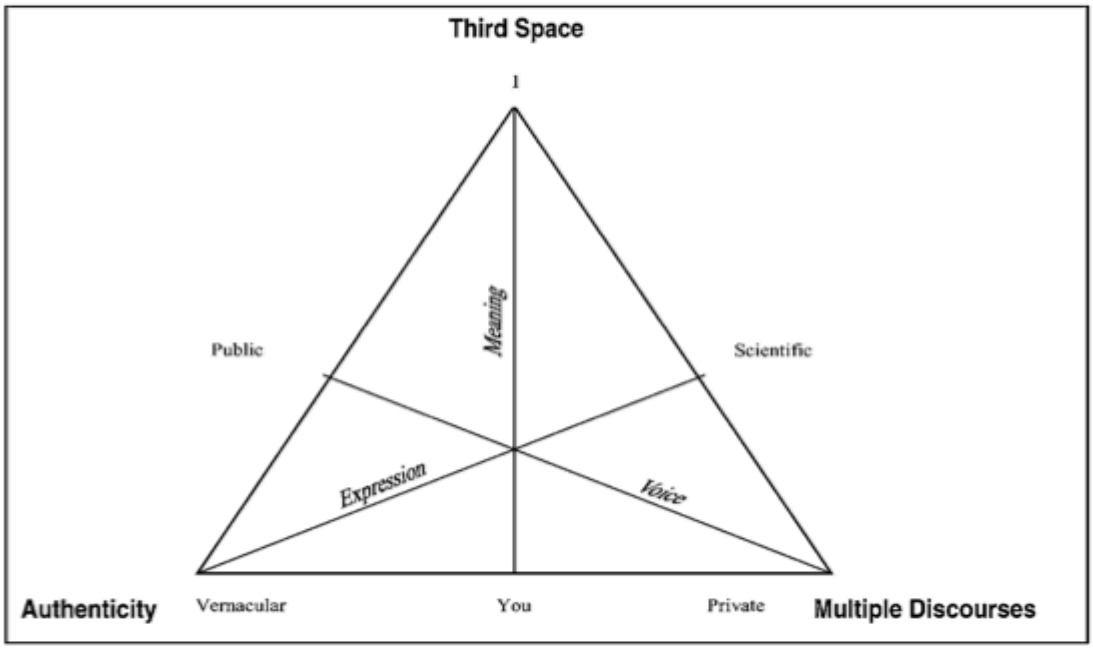


Figure 10. Theoretical Framework: Literacy and Language use in the science classroom (Wallace, 2004)

Implications

Improving American students' SL skills is clearly a national priority, but to reach this goal it is equally important to ensure that science teachers are in a position to prepare their students. This study found that secondary science teachers in one school district, although fairly scientific literate according to science education reform documents, are not able to develop student assessment tools that reflect the same SL goals. The implications, therefore, for policy makers, practitioners/administrators, and researchers are both pragmatic and theoretical.

Education policy. Colorado teachers are acutely aware of the state policy (SB191) that determines teacher security based on evaluations that reflect student performance on

standardized and district assessments. This policy is driving classroom practice, which is troubling when studies, such as mine, indicate that teacher-developed assessments designed to measure that effectiveness are misaligned with SL reform goals (Figure 8). This could, at a minimum, impede the attainment of policy related to increasing the SL of HS graduates, and in particular will be reflected in district assessments. Teachers are not educated in psychometrics and do not necessarily know how to develop well-constructed assessments. Southerland, Sowell, and Enderle (2011) interviewed 18 practicing teachers and discovered one area of science teachers' pedagogical discontentment lies in their ability to assess science learning. Teachers expressed low self efficacy with their abilities to carry out both traditional and alternative measures (the means of assessment), while at the same time struggled with how to interpret student performance on these assessments (the ends of assessment) (Southerland et al. (2011).

When 50% of a teacher's effectiveness rating is based on whether student learning objectives have been met, it implies that teachers have high quality assessments accessible to student measure. It behooves administrators and policy makers to reconsider policies that encourage teachers to develop poor quality content assessments (e.g., focus on recall of facts as opposed to application of scientific reasoning). Alternatively, if teachers recognize the importance of creating high quality assessments that measure SL but do not feel they have the skills to do so and if they do not receive the support, they may leave the science teaching profession altogether. This may mean that only teachers who are either not scientifically literate or those who are comfortable assessing (and teaching) students at a low SL level will be the ones who remain in the classrooms. If this occurs, national reform efforts to increase American students' SL levels are less likely to be met.

Professional development (PD). Teachers need professional development in SL (both content and practice) and assessment development. PD providers must acknowledge that a teacher's time is at a premium and that many teachers are skeptical of needing PD at all, thus getting teachers to the table is the first hurdle to overcome (Loucks-Horsley, Stiles, Mundry, Love, & Hewson, 2010). Understanding the conceptual change model is also essential for PD providers since we know that in order for teachers to adopt new conceptions they must be dissatisfied with their current views (Posner et al., 1982; Hewson, 1992). In addition, PD can and should be grounded within the context of teacher's practice, not separate. Two areas in particular for PD to highlight include: Science Literacy PCK and Assessment PCK.

Enhancing teachers PCK of science literacy. PD providers must first address teacher perceptions of SL because, as this study demonstrates, teachers will arrive at PD workshops with different ideas. Teacher PD on scientific inquiry is often insufficient in promoting conceptual change (Blanchard, Southerland, & Granger, 2009). Blanchard et al. (2009) found that only teachers who already had a strong understanding of the relationship between educational theory and practice benefitted from six-week long Research Experience for Teachers PD program intended to increase understanding of inquiry. Sometimes teachers are cognizant of the misalignment between their perceptions of SL and their practices; however, need guidance in modifying their teaching (Tobin, Briscoe, & Holman, 1990).

PD providers must challenge teachers to identify their own misconceptions related to SL; only then can teachers move into the shared *3rd Space* of meaning-making (Wallace, 2004) of SL and how to help students increase their understanding of it and then provide support during reform practice. Teacher perceptions indicated that many teachers identify their students' abilities to read and write scientifically / technically as SL. Because Norris and Phillips (2003)

argued that reading and writing about the content of science is fundamental to science literacy and should be taken seriously, this should be explored in SL PD. Knowing how to interpret and implement lessons on reading, writing, and meaning making of "non-content" vocabulary (e.g. 'Gather and analyze data related to…' or 'Use evidence to model…') identified in science content standards must be a priority for PD workshops since teachers do not share a common understanding of what these are. Teachers have reported that their knowledge and skills grew and their practice changed when they received content focused and learner-centered PD (Darling-Hammond & Richardson, 2009). Hence, PD that helps teachers explicitly explore the sub-constructs of SL is warranted and follow-up studies that measure student SL outcomes are needed.

The SL construct of scientific reasoning includes evidence-based reasoning (CER framework), quantitative reasoning, and argumentation. Despite the attention that argumentation has been given in science education reform literature, research has not shown that teachers regularly address CER in their science classrooms (Cavagnetto, 2010). Carlson (1991, in Sampson and Blanchard, 2012) argued that some teachers might not encourage students to evaluate alternative viewpoints or engage in argumentation because they believe science is a discrete body of knowledge to be learned. Kang et al., (2008) recommends that what "needs to be promoted in teacher PD is that connecting activities to scientific knowledge and communicating through argumentation, are just as important in inquiry activities as collecting data and formulating explanations (p. 350)." Hence, teachers need explicit instruction and PD on scientific reasoning (i.e., exploring when inductive and deductive reasoning strategies are used) and scientific argumentation.

Erduran et al. (2004) reported on the developments of applying Toulmin's (1958)

Argument Pattern to discourse analysis in science classrooms. They introduced two methodologies that can be useful in quantifying the usefulness of teachers' implementation of argumentation into their curriculum. They noted the challenges involved for teachers in teasing out what counts as claim, data, warrant, and backing during the analysis of classroom-based verbal data. In fact, I also encountered similar difficulties in my own content analysis of teacher open responses. One more sentence of clarification Erduran et al.'s (2004) evidence suggested that as teachers' skills improved with teaching and engaging students in argumentation discourse, students' argumentation abilities would follow.

The NRC (2012) recommended that students immerse themselves in the practices of scientists and engineers in order to appreciate the nature of how scientific knowledge has been generated within social contexts (i.e., it is necessary for scientists to share and evaluate each others' claims and data before they are accepted). For teachers to be able to do this, they must have strong pedagogical content knowledge (PCK). McNeill and Knight (2013) focused teacher PD on their PCK of argumentation as a means to successfully integrate the science and engineering practices outlined in the NGSSs. Four themes emerged from their study including that: (1) Teachers' ability to identify evidence in student writing was strong, and their understanding of claim and reasoning improved, but, within classroom practice, teaching "reasoning" was still challenging for teachers; (2) For classroom discussion, teachers had a limited understanding of argumentation; (3) Teachers found that designing questions for argumentation was challenging but necessary; and (4) Elementary teachers connected argumentation to other disciplines, whereas high school teachers focused more on science content. Hence, teachers can benefit from PD on SL reasoning and argumentation; however, they

also need to participate in scientific argumentation as learners to be better prepared to facilitate students' argumentative discourse.

Likewise, Crippen's (2012) study *Argument as Professional Development: Impacting Teacher Knowledge and Beliefs about Science* involving 42 secondary science teachers, confirmed that argumentation is a successful method for increasing science teacher content knowledge. However, Crippen's study reported that teachers had a lack of trust for using their own data when making claims and were more apt to use the Internet and cite an outside source as evidence. The premise that teachers are not comfortable with their own knowledge construction calls into question their abilities to lead students in their own inquiries and their understanding of the purpose of scientific work.

Almost half of the participating teachers in this study had not read the NGSSs, nor did they feel like they had enough information to comment on how they differed from current and previous standards. The scientific community has already dealt with the problematic conflation and/or equating of the abilities of Scientific Inquiry with the Understandings of Science inquiry. With additional skills such as those outlined in the practices of Science and Engineering, teachers must receive PD on the goals of the NGSS. Of course if "Engaging students in the practices is not sufficient for SL" (Achieve, 2013a, p. 16) as the NGSS states, it will certainly not be sufficient for teacher PD. Achieve (2013a) recommends that students stand back and reflect on how these practices have contributed to their own development and to the accumulation of scientific knowledge and engineering accomplishments over the ages. Therefore, teachers also can benefit from moving beyond just scientific inquiry PD to opportunities to engage in synthesis and evaluation of past scientific findings that prepare them to develop scientific arguments.

Enhancing Teacher's PCK of Assessment. Historically, teachers have been ill prepared to meet the assessment demands of today's classrooms (Stiggins, 2001), thus secondary science teachers charged with developing high-stakes assessments need PD on how to do so. According to McMillan (2003) many teachers' assessment decision-making is characterized by the conflict between their internal beliefs and values about assessment and the external demands of high-stakes assessment policies. Stiggins, (2001) highlighted five assessment standards that can support teachers in their assessment decision-making and enhance their assessment PCK. These include: a) begin with having clear learning targets; b) identify the instructional purpose for the assessment; c) match the assessment method with the intended target; d) make sure each assessment is representative of students performance; and e) eliminate bias in test questions and administration procedures (Stiggins, 2001). Teachers need to be given the professional time and guidance to improve their performance on assessment standards; which can lead to an improvement in their overall assessment literacy.

Abell and Siegal (2011) presented a coding dictionary for PCK assessment categories that resulted from their analysis of pre-service teachers from two science methods courses through their first year teaching. They constructed a model of *science teacher assessment literacy*, which centers on teachers' views of learning, including the assessment values and principles that guide them. The values and principles reflect four categories of knowledge about assessment – purpose, content, strategies, and interpretation/resulting actions (Abell & Siegal, 2011). This model can easily be implemented in PD around how to assess SL. Another useful PD approach is to help teachers understand how alignment between data and content maps is important (Porter et al., 2007). The use of Porter's Alignment Index has been shown to be effective for increasing the alignment between the expected and the implemented curriculum. By empowering teachers to

learn how to calculate alignment, Porter et al. (2007) found a stronger alignment between teacher-constructed assessments and their learning objectives.

Since the inception of the NSES in 1996 the need for resources to teach scientific inquiry as well as the access to appropriate and meaningful assessments has been documented (Lederman et al., 2002). Lederman et al.'s (2002) Views of the Nature of Science of VNOS questionnaire is a valid and reliable instrument for determining student understanding of and essential SL skill, yet teachers are often hesitant to use an open response instrument because they will need to code the responses (personal observation). Coding such responses assumes that teachers have a strong command of the possible answers and what is scientifically acceptable. As a result, multiple-choice SL assessments, like the TOSLS, are promising resources that teachers can use to identify strategies to assess SL with forced response items because answer keys are usually available.

Implications for research. Future research is warranted to better understand teacher perceptions of SL, including fundamental science literacy. Fundamental literacy skills are foundational to developing SL assessments because the reading and interpretation of "non-content" words in the standards are just as important for teachers to agree upon as are the content. In other words, do teachers not know that they are settling for lower assessment skills? Collaborations between English language education and science education researchers are needed to uncover how science teachers use and assess student science knowledge and skills through language. Parallel studies of student SL gains (through mixed methods such as performance on instruments like the TSOLS, as well as follow-up interviews) would allow us to determine the relationship between teacher SL levels and their respective students. In fact, the district Director for Research and Assessment was supportive of this study because of plans for

conducting a follow up study such as this. Because the response rate of teachers was too low to examine the relationship between teacher SL level and student performance on state standardized tests, more efforts will need to be made to recruit more teacher participants.

Future studies on the process of teacher assessment development can also be informative. Teacher focus groups on the process of developing common assessments will allow us to determine if teachers can distinguish between "low quality" and "high quality" assessments, and if they care or not. One would predict that teachers might choose to develop assessments that center on fact recall because of the consequences of students performance on their own job security. However, are teachers making a conscious decision to do this or not? This cannot be answered without follow up studies. If districts set up incentive programs that compare student performance at different schools within the district, future studies may use Game Theory to better understand the decisions and choices teachers make. In other words, if all teachers choose to develop low quality assessments, then no one "loses," but if some teachers choose to create more challenging assessments, it will be to the detriment of other teachers. Do teachers make such decisions using this information or reasoning?

More straightforward studies on assessment can include studying the impact of different PD models on teacher outcomes. Do teachers benefit from engaging in authentic scientific inquiry and discourse (such as through RETs: research experience for teachers or Teacher-In-Residence programs)? Is it sufficient to tell teachers how to create better assessments and then follow up with them throughout the school year? Do teachers create better assessments when they can engage in lesson study and observe each other's classrooms? In any event, the follow-up research questions are endless but worthy of study if we hope to increase both secondary science teacher and student scientific literacy.

Limitations

As with any study there were limitations that are important to identify. The major limitations of this study were the small sample size and the narrow demographic region selected for this study. This will affect my ability to generalize and apply results to other districts in the state and across the nation. Despite methodical and regular efforts to recruit teacher participants, it was difficult to recruit the target number of teachers. It may have been because they did not necessarily see the value of the district or university identifying their weaknesses in order to identify future PD or because they started the TOSLS and were overwhelmed with the length of the test, or possibly they did not feel prepared to complete it. Sixteen teachers began the survey and only completed the demographic data another 11 started the TOSLS, but never completed the instrument. Hence, another limitation may be that only those who felt competent and confident were the ones who completed the SL instrument.

Another limitation of this study was that the survey results were based solely on participating teachers' self- reported frequencies of both teaching and assessing scientific literacy skills; this may have influenced results of research question two (determining the difference between teachers ranking of the frequency of importance versus assessment of scientific literacy skills). In addition, some teachers expressed their concern (verbally and not through the survey) that they are not experts in test development but, more importantly, that their students' performance on these assessments could "hurt" the teachers. As with any perception survey, the researcher is constrained to how the participants chose to respond at the moment of completion.

Despite the localized data collection, these study findings should be of interest to educators and administrators around the state and nation. The dilemma of how to assess scientific literacy skills across districts and to assess teachers on their effectiveness teaching

scientific literacy is timely and relevant. There are few studies that have addressed this issue in part because the NGSSs were only released in March 2013 and teacher effectiveness measures based on policy are in the early stages of enactment or are still being developed in some areas.

Conclusion

Conceptual change theorists Posner, Strike, Hewson, and Gertzog (1982) argued that there are two phases of conceptual change: assimilation and accommodation. Assimilation involves the explanation of new learning using existing conceptual understandings, although often times the learners' current concepts are inadequate and a more radical form of conceptual change is needed, which is accommodation (Posner et al., 1982). The conditions necessary for accommodation to occur include 1) dissatisfaction with existing conceptions; 2) the new conception must be intelligible; 3) the new conception must be plausible and reinforced by prior knowledge; and 4) the new concept must have the potential to be expand and open up new areas of "fruitful research" (Posner et al., 1982).

This study has shown that teachers' current perception of SL may be inadequate and their abilities to assess SL sub-constructs at a collective level are also not evident, thus it behooves education reformers and policy makers to continue to strive for common, intelligible, language and assessments devoted to achieving SL in our high school graduates. The time has come to stop just talking about reform and work collaboratively towards teachers' accommodation of these goals.

REFERENCES

Abd-El-Khalick, F., & Lederman, N. G. (2000). The influence of history of science courses on students' views of nature of science. *Journal of Research in Science Teaching, 37*, 1057-1095.

Abell, S. K. (2007). Research on science teacher knowledge. In Abell, S. & Lederman, N. (Eds.) *Handbook of Research on Science Education* (pp. 1105-1149). New York, NY: Routledge.

Abell, S., & Siegal, M. (2011). Assessment literacy: What science teachers need to know and be able to do. In Corrigan, D., Dillon, J., & Gunstone, R. (eds.), *The Professional Knowledge Base of Science Teaching* (pp. 205-221). London: Springer Science.

Abott, (ND). The elementary and secondary education act of 1965. Retrieved from http://www.socialwelfarehistory.com/events/elementary-and-secondary-education-act-of-1965/

Acar, O., Turkmen, L., & Roychoudhury, A. (2010). Student difficulties in socio-scientific argumentation and decision-making research findings: Crossing the borders of two research lines. *International Journal of Science Education, 32*(9), 1191-1206.

Achieve, Inc. (2013a). Next Generation Science Standards. Retrieved from www.nextgenscience.org

Achieve, Inc. (2013b). APPENDIX F – Science and Engineering Practices in the NGSS. Retrieved from http://www.nextgenscience.org/sites/ngss/files/Appendix%20F%20%20Science%20and%20Engineering%20Practices%20in%20the%20NGSS%20-%20FINAL%20060513.pdf

Allchin, D. (2011). Evaluating knowledge of the nature of (whole) science. *Science Education, 95,* 518-542.

Alvarado, A. (2010). The interaction of MI environmental education curriculum, science teachers' PCK, and environmental action competence. (Doctoral dissertation), UMI Number: 3435267.

American Association for the Advancement of Science. (1990). *Science for all Americans.* Washington DC: Author.

American Association for the Advancement of Science. (1993). *Benchmarks for scientific literacy*. Washington DC: Author.

Anderson, K. (2012). Science education and test-based accountability: Reviewing their relationship and exploring implications for future policy. *Science Education 96*(1), 104-129.

Arons, A. B. (1983). Achieving wider scientific literacy. *Daedalus*. *112*(2), 91-122.

Association of American Colleges and Universities, (2012). VALUE: Valid assessment of learning in undergraduate education. Quantitative literacy VALUE rubric. Retrieved from http://www.aacu.org/value/rubrics/QuantitativeLiteracy.cfm

Balgopal, M. M. and Wallace, A. M. (2009). Dilemmas and decisions: The use of guided writing to increase ecological literacy of elementary education majors. *J. Environmental Education, 40*(3), 13-26.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice Hall.

Bandura, A. (1997). *Social learning theory*. Englewood Cliffs, NJ: Prentice Hall.

Banilower, E. R., Smith, P., Weiss, I., Malzahn, K., Campbell, K., Weis, A. (2013). Report of the 2012 national survey of science and mathematics education. Horizon Research. Retrieved from http://www.horizon-research.com/2012nssme/research-products/reports/technical-report/

Bell, R. L., Lederman, N. G., & Abd-El-Khalick, F. (2000). Developing and acting upon one's conception of the nature of science: A follow-up study. *Journal of Research in Science Teaching*, *37*(6), 563-581.

Bhabha, H. (1994). *The location of culture.* New York: Routledge.

Blanchard, M. R., Southerland, S. A., & Granger, E. M. (2009). No silver bullet for inquiry: Making sense of teacher change following an inquiry-based research experience for teacher. *Science Education, 93*(2), 322-360.

Bloom, B. S. (1967). *Taxonomy of educational objectives: The classification on educational goals handbook I, cognitive domain.* New York: David McKay.

Briggs, D. C. (2013). Teacher evaluation as Trojan house: The case for teacher-developed assessments. *Measurement: Interdisciplinary Research and Perspectives. 11*(1-2) 24-29.

Brown, N. J. S., Furtak, E. M., Timms, M., Nagashima, S. O., & Wilson, M. (2010). The evidence-based reasoning framework: Assessing scientific reasoning. *Educational Assessment*, *15*, 123-141.

Busch, C., De Maret, P. S., Flynn, T., Kellum, R., Le, S., Meyers, B., . . . Palmquist, M. (1994-2000). Content Analysis. Writing @CSU. Colorado State University. Retrieved from writing.colostate.edu/guides/guide.cfm?guideid=61.

Bybee, R., Fensham, P., & Laurie, R. (2009). Scientific literacy and contexts in PISA 2006 science. *Journal of Research in Science Teaching, 46*(8), 862-864.

Bybee, R. W., Powell, J. C., & Trowbridge, L. W. (2008). *Teaching secondary school science: Strategies for developing scientific literacy.* Upper Saddle River, NJ: Pearson Publishers.

Cadbury, D. (2006). *Space race: The epic battle between America and the Soviet Union for dominion of space*. New York: Harper Collins.

Carr, R., Bennett, L., & Strobel, J. (2012). Engineering in the K-12 STEM standards of the 50 U.S. States: An analysis of presence and extent. *Journal of Engineering Education*, *101*(3), 539-564.

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

Chen, Y. C. (2011). *Examining the integration of talk and writing for student knowledge construction through argumentation* (Doctoral dissertation). Retrieved from University of Iowa database: 2011.http://ir.uiowa.edu/etd/1129

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*, 37-46.

College Board. (2009). Science: College board standards for college success. Retrieved from http://professionals.collegeboard.com/profdownload/cbscs-science-standards-2009.pdf

CDE. (2009). Colorado Academic Standards: Science. Retrieved from http://www.cde.state.co.us/CoScience/StateStandards.asp

CDE. (2010). Educator effectiveness: A Colorado priority. Retrieved from http://www.cde.state.co.us/EducatorEffectiveness/index.asp

CDE. (2012a). Determining high-quality content assessments. Retrieved from Colorado content collaborative webinar: Content review tool web site: http://www.cde.state.co.us/contentcollaboratives/ResourceBank.asp

CDE. (2012b). Colorado education statistics. Retrieved from http://www.cde.state.co.us/index_stats.htm

CDE. (2013). State model evaluation system for teachers. Retrieved from http://www.cde.state.co.us/educatoreffectiveness/smes-teacher

CDE. (2014). CDE Resource bank. Retrieved from http://www.coloradoplc.org/assessment

Committee of Ten. (1894). Report of the committee of ten on secondary school studies. Retrieved from http://ia600409.us.archive.org/25/items/cu31924032709960/cu31924032709960.pdf

CCSSI. (2013a). English language arts standards, writing: Grade 9-10. Retrieved from http://www.corestandards.org/ELA-Literacy/W/9-10

CCSSI. (2013b). Implementing the common core state standards. Retrieved from http://www.corestandards.org/

Contino, J. (2013). A case study of the alignment between curriculum and assessment in the New York State earth science standards-based system. *Journal of Science Education and Technology, 22*, 62-72.

Crippen, K. J. (2012). Argument as professional development: Impacting teacher knowledge and beliefs about science. *Journal of Research in Science Teacher Education, 23*, 847–866.

Crotty, M. (1998). *The foundations of social research: Meaning and perspective in the research process.* Thousand Oaks, CA: Sage Publications.

Curie M. (ND). Marie Curie Quotes. Retrieved from http://womenshistory.about.com/od/quotes/a/marie_curie.htm

Danielson, C. (2007). *Enhancing professional practice: A framework for teaching.* Alexandria, VA: Association for Supervision and Curriculum Development.

Darling-Hammond, L. (1986). Teaching Knowledge: How Do We Test It? *American Educator: The Professional Journal of the American Federation of Teachers*, *10*(3), 18-21.

Darling-Hammond, L. (1994). Performance-based assessment and educational equity. *Harvard Educational Review, 64*(1), 5-31.

Darling-Hammond, L., & Pecheone, R. (March, 2010). *Developing an internationally comparable balanced assessment system that supports high-quality learning*. Proceedings of the National Conference on Next Generation Assessment Systems, Princeton, NJ.

Darling-Hammond, L., & Richardson, N. (2009). Research review/teacher learning: What matters. *Educational leadership*, *66*(5), 46-53.

Darling-Hammond, L., Amrein-Beardsley, A., Haertel E., & Rothstein, J. (2012). Evaluating teacher effectiveness. *Phi-Delta Kappan*, *93*(6), 8-15.

Demir, A. & Abell, S.K. (2010). Views of inquiry: Mismatches between vies of science education faculty and students of an alternative certification program. *Journal of Research in Science Teaching, 47* (6), 716-741.

Dilworth, M. E. (2013). National Board for Professional Teaching Standards. In Hattie, J. & Anderman, E. (eds) *International Guide to Student Achievement* (pp. 224-227). New York: Routledge.

Divine, R. A. (1993). *The Sputnik Challenge: Eisenhower's Response to the Soviet Satellite.* New York: Oxford University Press.

Downing, S. M., & Haladyna, T. M. (Eds). (2006). *Handbook of test development*. Mahwah, NJ: Lawrence Erbaum Associates.

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

Duschl, R. A. (2003). Assessment of inquiry. In J. M. Atkin & J. Coffey (eds.), *Everyday assessment in the science classroom* (pp. 41–59). Arlington, VA: NSTA Press.

Duschl, R. (May-June, 2012). *Quantitative reasoning and mathematical modeling in STEM: A commentary.* Paper from an International STEM Research Symposium: QR in Mathematics and Science Education, Savanna, GA. Retrieved from http://coe.georgiasouthern.edu/QR/QR_readings.html

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

Evagorou, M., Jimenez-Aleixandre, M. P., & Osborne, J. (2012). Should we kill the grey squirrels? A study exploring students' justifications and decision-making. *International Journal of Science Education, 34*(3), 401-428.

Evagorou, M. & Dillon, J. (2011) Argumentation in the teaching of science. In D. Corrigan, Dillon, J., & Gunston, R. (eds.), *The professional knowledge base of science teaching.* Dordrecht: Springer.

Evans, J. R., & Mathur, A. (2005). The value of online surveys. *Internet Research,* 15(2), 196-219.

Garber, S. (2007). Sputnik: Sputnik and the dawn of the space age. Retrieved from http://history.nasa.gov/sputnik/

Gates Foundation. (2013). Ensuring fair and reliable measure of effective teaching: Culminating findings from the MET project's three-year study. Retrieved from www.metproject.org/

Gess-Newsome, J., Southerland, S., Johnston, A., & Woodbury, S. (2003). Educational reform, personal practical theories, and dissatisfaction: The anatomy of change in college science teaching. *American Educational Research Journal*, *40*(3), 731-767.

Glaser, E. M. (1941). An experiment in the development of critical thinking. *The Teacher's College Record, 43*(5), 409-410.

Glaser, B. G., & Strauss, A. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago: Aldine.

Gliem, J. A., & Gliem, R. R. (2003). Calculating, interpreting, and reporting Cronbach's alpha reliability coefficient for Likert-type scales. Midwest Research-to-Practice Conference in Adult, Continuing, and Community Education.

Gliner, J., Morgan, G., & Leech, N. (2009). *Research methods in applied settings: An integrated approach to design and analysis, (2nd ed.)*. New York, NY: Routledge, Taylor and Francis Group.

Goe, L., Bell, C., & Little, O. (2008). Approaches to evaluating teacher effectiveness: A research synthesis. National Comprehensive Center for Teacher Quality: Washington, DC.

Goe, L., & Holdheide, L. (2011). Measuring teachers' contributions to student learning growth for non-tested grades and subjects. National Comprehensive Center for Teacher Quality: Washington, DC.

Goehrig, G. H. & Luft, J. A. (2007). Constraints experienced by beginning secondary science teachers in implementing scientific inquiry lessons. *International Journal of Science Education, 26*(1), 3-24.

Goldhaber, D. & Hansens, S. (2010). Teacher Tenure Decisions. *American Economic Review: Papers and Proceedings,* 250-255.

Gormally, K., Brickman, P., & Lutz, M. (2012). Developing a test of scientific literacy skills (TOSLS): Measuring undergraduates' evaluation of scientific information and arguments. *CBE–Life Sciences Education*, *11*(4), 264-377.

Guba, E.G. (1981). Criteria for assessing the trustworthiness of naturalistic inquiries. *Educational Communication and Technology Journal, 29*(2), 75-91.

Halliday, M. A. K. & Martin, J. R. (1993). *Writing science: Literacy and discursive power.* London, U.K.: Falmer Press.

Harmon, A. (2011). It may be a sputnik moment, but science fairs are lagging. Retrieved from http://www.nytimes.com/2011/02/05/us/05science.html?_r=0

Herman, J.L., Heritage, M., & Goldschmidt, P. (2011). *National Center for Research on Evaluation, Standards, & Student Testing (extended version).* Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Hewson, P. W. (June, 1992). Conceptual change in science teaching and teacher education. Proceedings from *National Center for Educational Research, Documentation, and Assessment*, Madrid, Spain.

Heubert, J., & Hauser, R. (Eds.) (1999). *High stakes: Testing for tracking, promotion, and graduation.* Washington DC: National Academy Press.

Johnson, C., Zhang, D., & Kahle, J. (2012). Effective science instruction: Impact on high-stakes assessment performance. *Association for Middle Level Education, 35*(9), 1-14.

Jones, M. G., & Carter, G. (2007). Science teacher attitudes and beliefs. In S. Abell & N. Lederman, *Handbook of Research on Science Education* (pp. 1067-1104). New York, NY: Routledge.

Kang, N-H, Orgill, M., & Crippen, K. J. (2008). Understanding teachers' conceptions of classroom inquiry with a teaching scenario instrument. *Journal of Science Teacher Education*, *19*, 337–354.

Kolsto, S. D. (2001). Scientific literacy for citizenship: Tools for dealing with the science dimension of controversial socioscientific issues. *Science Education, 85*, 291-310.

Kolsto, S. D. (2006). Patterns in students' argumentation confronted with a risk-focused socio-scientific issue. *International Journal of Science Education, 28*(14), 1689-1716.

Krippendorff, K. (1980). Content analysis: An introduction to its methodology. Thousand Oaks, CA: Sage Publications.

Krippendorff, K. (2004). Content analysis: An introduction to its methodology (2nd ed.). Thousand Oaks, CA: Sage Publications.

Kuhn, T. (1962). *The structure of scientific revolutions.* Chicago: University of Chicago Press.

Kuhn, D. (1993). Science as argument: Implications for teaching and learning scientific thinking. *Science Education*, *77*(3), 310-337.

Kurz, A., Elliott, S. N., Wehby, J. H., & Smithson, J. L. (2010). Alignment of the intended, planned, and enacted curriculum in general and special education and its relation to student achievement. *The Journal of Special Education*, *44*(3), 131-145.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *biometrics*, *33*(1), 159-174.

Larkin, D. B., Seyforth, S. C., & Lasky, H. J. (2009). Implementing and sustaining science curriculum reform: A study of leadership practices among teachers within a high school science department. *Journal of Research in Science Teaching, 46*(7), 813-835.

Laugksch, R.C. (2000). Scientific literacy: A conceptual overview. *Science Education, 84*, 71-94.

Lawson, A. (2010). *Teaching inquiry science in middle and secondary schools*. Thousand Oaks, CA: Sage Publications.

Lederman, N.G. (1992). Students' and teachers' conceptions of the nature of science: A review of the research. *Journal of Research in Science Teaching, 29*, 331–359.

Lederman, N. G., Abd-El-Khalik, F. R., & Schwartz, R. S. (2002). Views of nature of science questionnaire: Toward valid and meaningful assessment of learners' conceptions of nature of science. *Journal of Research in Science Teaching, 39*(6), 497-521.

Lee, E., Brown, M., Luft, J., & Roehrig, G. (2007). Assessing beginning science teachers' PCK: Pilot year results. *School Science and Mathematics*, *107*(2), 52-60.

Liu, X., Zhang, B., Liang, L., Fulmer, G., Kim, B., & Yuan, H. (2009). Alignment between the physics content standard and the standardized test: A comparison among the United States-New York state, Singapore, and China-Jiangsu. *Science Education. 93*(5), 777-797.

Liu, X. & Fulmer, G. (2008). Alignment between the science curriculum and assessment in selected NY state regents exams. *Journal of Science Education and Technology*, *17*, 373-383.

LoGiudici, R., & Ende, F. (July, 2010). Science sampler: Teaching toward a more scientifically literate society. Science Scope

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

Loucks-Horsley, S., Stiles, K. E., Mundry, S., Love, N., & Hewson, P. W. (2010). *Designing professional development for teachers of science and mathematics* (3rd ed.). Thousand Oaks, CA: Corwin Press.

Loucks-Horsley, S., & Matsumoto, C. (1999). Research on professional development for teachers of mathematics and science: The state of the scene. *School Science and Mathematics*, *99*, 258–271.

Magnusson, S., Krajcik, J., & Borko, H. (1999). Nature, sources and development of pedagogical content knowledge for science teaching, in J. Gess-Newsome & N. Lederman (Eds), PCK and Education, (pp. 95-132). The Netherlands: Kluwer Academic Publishers.

McLeod, S. (2008). Likert scale. *Simply Psychology*.

McMillan, J. (2003). Understanding and improving teachers' classroom assessment decision making: Implications for theory and practice. *Educational Measurement: Issues and Practice, 22*: 34-43.

McNeill, K. L., & Knight, A. M. (2013). Teachers' pedagogical content knowledge of scientific argumentation: The impact of professional development on K-12 teachers. *Science Education, 97*, 936-972.

McNeill, K. L., & Krajcik, J. S. (2009). Synergy between teacher practices and curricular scaffolds to support students in using domain-specific and domain-general knowledge in writing arguments to explain phenomena. *The Journal of the Learning Sciences, 18*, 416-446.

McNeill, K. L., & Krajcik, J. S. (2012). *Supporting grade 5-8 students in constructing explanations in science: The claim, evidence, and reasoning framework for talk and writing*. Boston, MA: Pearson Education.

Mertler, C. A., & Campbell, C. (April, 2005). Measuring teachers' knowledge and application of classroom assessment concepts: Development of the assessment literacy inventory. In paper presented at the annual meeting of the American Educational Research Association, Montreal, Quebec, Canada.

Metz, S. (May, 2006). Editor's corner—Science literacy: Then and now. *The Science Teacher.*

Midgley, R. (September, 2009). Evidence-based reasoning. Retrieved from www.indepthlearning.org/2009/09/evidence-based-reasoning/

Moss, P.A., Pullin, D. C., Gee, J. P., Haertel, E. H., & Jones Young, L. (eds.) (2008). *Assessment, equity, and opportunity to learn.* New York: Cambridge University Press.

National Governors Association. (2010). *Common core state standards.* National Governors Association Center for Best Practices, Council of Chief State School Officers, Washington D.C.

NRC. (1996). *National science education standards*. Washington, DC: National Academy Press.

NRC. (2000a). *How people learn: Brain, mind, experience, and school.* Washington, DC: National Academy Press.

NRC. (2000b). Inquiry and the national science education standards: A guide for teaching and learning. Washington, DC: National Academy Press.

NRC. (2001). *Classroom assessment and the national science education standards.* Washington, DC: National Academy Press.

NRC. (2007). *Taking science to school: Learning and teaching science in Grades K-8.* Washington, DC: National Academy Press.

NRC. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas.* Washington, DC: National Academy Press.

NSTA. (2010). Teaching science and technology in the context of societal and personal issues. NSTA Position Statement. Retrieved from http://www.nsta.org/about/positions/societalpersonalissues.aspx

Nicolaidou, I., Kyza, E., Terzian, F., Hadjichambis, A., & Kafouris, D. (2011). A framework for scaffolding students' assessment of the credibility of evidence. *Journal of Research in Science Teaching, 48*(7), 711-744.

Neuendorf, K. (2002). *The content analysis guidebook*. Thousand Oaks, CA: Sage Publications.

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

OECD. (1999). Measuring student knowledge and skills: A new framework for assessment. Paris: Author.

OECD. (2009). PISA 2009 Assessment framework: Key competencies in reading mathematics and science. Paris: Author.

Orpwood, G. (2001). The role of assessment in science curriculum reform. *Assessment in Education, 8*(2), *135-151.*

Osborne, J., & Patterson, A. (2011). Scientific argument and explanation: A necessary distinction? *Science Education*, *95*(4), 627-638.

Parcell T., & Dufur, M. (2001). Capital at home and at school: Effects on student achievement. *Social Forces*, *79*(3), 881-912.

Park, S., & Oliver, J. S. (2008). Revisiting the conceptualization of pedagogical content knowledge: PCK as a conceptual tool to understand teachers as professionals. *Research in Science Education, 38*, 261-284.

Partnership for 21st Century Skills (2002). *About us*. Retrieved from http://www.21stcenturyskills.org/index.php?option=com_content&task=view&id=188&Itemid=110

Pella, M. O., O'Hearn, G. T., & Gale, C. W. (1966). Referents to scientific literacy. *Journal of Research in Science Teaching, 4*, 199-208.

Poincaré, J. H. (1917). La science and l'hypothese. Paris: E. Flamarion.

Popham, J. (1999). Why standardized tests don't measure educational quality. *Educational Leadership*, *56*(6), 8-15.

Porter, A. C., (2002). Measuring the content of instruction: Uses in research and practice. *Educational researcher, 31*(7), 3-14.

Porter, A. C., Smithson, J., Blank, R., & Zeidner, T. (2007). Alignment as a teacher variable. *Applied Measurement in Education*, *20*(1), 27-51.

Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education, 66*(2), 211-227.

Poudre School District. (2010). Standards and assessments: $1 million of mill levy resources for instructional support for teachers. Retrieved from http://www.psdschools.org/node/5015

Poudre School District. (2013). Poudre School District student learning objectives guide. Retrieved from https://www.psdschools.org/webfm/6210/view

Quellmalz, E., Timms, M., Silberglitt, M., & Buckley, B. (2012). Science assessments for all: Integrating science simulations into balanced state science assessment systems. *Journal of Research in Science Teaching, 49*(3), *363-393.*

Roberts, D. A. (2007). Scientific literacy/Science literacy. In S. Abell & N. Lederman, *Handbook of Research on Science Education* (pp. 729-780). New York, NY: Routledge.

Roberts, C. M. (2010). *The dissertation journey: A practical and comprehensive guide to planning, writing, and defending your dissertation.* Thousand Oaks, CA: Sage Publications.

Sadler, T. D. (2004). Informal reasoning regarding socioscientific issues: A critical review of research. *Journal of Research in Science Teaching, 41*(5), 513-536.

Sadler, T. D., & Zeidler, D. L. (2005). Patterns of informal reasoning in the context of socioscientific decision-making. *Journal of Research in Science Teaching, 42*(1), 112-138.

Sadler, T. D., & Zeidler, D. L. (2009). Scientific literacy, PISA, and socioscientific discourse: Assessment for progressive aims of science education. *Journal of Research in Science Teaching, 46*(8), 909-921.

Sampson, V., & Blanchard, M. (2012) Science teachers and scientific argumentation: Trends in views and practice. *Journal of Research in Science Teaching, 49*(9), 1122-1148.

Sampson, V., & Clark, D. (2009). The impact of collaboration on the outcomes of scientific argumentation. *Science Education*, *93*(3), 448-484.

Schreir, M. (2012). *Qualitative content analysis in practice*. London: Sage Publications.

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, *15*(2), 4-14.

Scott, E.C. (2004). *Evolution vs. creationism: An introduction.* Berkeley: University of California Press.

Shen, B. S. (1975). Science literacy and the public understanding of science. *Communication of scientific information*, 44-52.

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

Shepard, L. (2007). A brief history of accountability testing, 1965–2007. In K. Ryan & L. Shepard (Eds.), *The Future of Test-Based Educational Accountability.* New York, NY: Routledge.

Showalter, V. M. (1974). What is unified science education? Program objectives and scientific literacy. *Prism II*, *2*, 3-4.

Siegel, M. A., & Wissehr, C. (2011). Preparing for the plunge: Preservice teachers' assessment literacy. *Journal of Science Teacher Education*, *22*(4), 371-391.

Snow, C. P. (1959). *The two cultures and the scientific revolution.* Cambridge: Cambridge University Press.

Songer, N. B., Lee, H. S. & Kam, R. (2002), Technology-rich inquiry science in urban classrooms: What are the barriers to inquiry pedagogy? *Journal of Research in Science Teaching*, *39*, 128–150.

Steele, L., Hamilton, L., & Stecher, B. (2010). *Incorporating student performance measures into teacher evaluation systems*. Report sponsored by the Center for American Progress. Santa Monica, CA: Rand Education.

Smolin, L., (2008). Lee Smolin on science and democracy. Synopsis of a Ted talk retrieved from http://culture unplugged.com

Stemler, S. (2001). An overview of content analysis. *Practical Assessment, Research and Evaluation*, *7*(17). Retrieved from http://pareonline.net/getvn.asp?v=7&n=17

Stiggins, R. (1999). Assessment, student confidence, and school success. *Phi Delta Kappan*, *81*(3), 191-198.

Stiggins, R. (2001). The principal's leadership role in assessment. *NASSP Bulletin, 85*(621), 13-26.

Strauss, A. & Corbin, J. (1990). *Basics of Qualitative Research: Ground Theory, Procedures, and Techniques.* Newbury Park, CA: Sage Publications.

Southerland, S. A., Sowell, S., & Enderle, P. (2011). Science teachers' pedagogical discontentment: Its sources and potential for change. *Journal of Science Teacher Education, 22,* 437-457.

Tobin, K., Briscoe, C., & Holman, J.R. (1990). Overcoming constraints to effective elementary science teaching, *Science Education, 74,* 409-420.

Toulmin, S. (1958). *The Uses of Argument.* New York, NY: Cambridge University Press.

Trowbridge, L., Bybee, R., & Powell, J. C. (2008). *Teaching secondary school science: Strategies for developing scientific literacy*. Upper Saddle River, NJ: Pearson Publishers.

Understanding Science. (2012). University of California Museum of Paleontology. Retrieved from http://www.understandingscience.org

USDOE (2001). The elementary and secondary education act (The No Child Left Behind Act of 2001). Retrieved from http://www2.ed.gov/policy/elsec/leg/esea02/index.html

USDOE (2006). Proven methods: Highly qualified teachers for every child. Retrieved from http://www2.ed.gov/nclb/methods/teachers/stateplanfacts.html

USDOE (2009a). A blueprint for reform: The reauthorization of the elementary and secondary education act. Retrieved from http://www2.ed.gov/policy/elsec/leg/blueprint/publication_pg5.html#part5

USDOE (2009b). U.S. Department of education opens race to the top competition. Retrieved from http://www2.ed.gov/news/pressreleases/2009/11/11122009.html

Van Driel, J. H., Verloop, N., & De Vos, W. (1998). Developing science teachers' pedagogical content knowledge. *Journal of Research in Science Teaching*, *35*, 673–695.

Venville, G., & Dawson, V. (2010). The impact of a classroom intervention on grade 10 students' argumentation skills, informal reasoning, and conceptual understanding of science. *Journal of Research in Science Teaching*, *47*(8), 952-977.

Wagner, T. (2008). *The global achievement gap: Why even our best schools don't teach the new survival skills our children need—and what we can do about it.* New York, NY: Basic Books.

Wallace, C. S. (2004). Framing new research in science literacy and language use: Authenticity, multiple discourses, and the "Third Space." *Science Education, 88*(6), 901-914.

Wallace, C. S. (2012). Authoritarian science curriculum standards as barriers to teaching and learning: An interpretation of personal experience. *Science Education, 96(*2), 291-310.

Wang, H.A., & Marsh, D. D. (2002). Science instruction with a humanistic twist: Teachers' perception and practice in using the history of science in their classrooms. *Science & Education, 11*, 169-189.

Weber, R., P. (1990). *Basic content analysis: Quantitative applications in the social sciences*. Newbury Park, CA: Sage Publications.

Whilden, B. E. (2010). The elementary and secondary education act: A Primer on reauthorization in 2010. Retrieved from http://www.congressweb.com/aascu/docfiles/ESEA%20PRIMER%@)FINAL.pdf

Wilson, S. M., & Wineberg, S.S. (1993). Wrinkles in time and place: Using performance assessments to understand the knowledge of history teachers. *American Educational Research Journal, 30*(4), 729-769.

Windschitl, M. (2003). Inquiry projects in science teacher education: What can investigative experiences reveal about teacher thinking and eventual classroom practice? *Science Education, 87*(1), 112-143.

Woodbury, S. & Gess-Newsome, J. (2002). Overcoming the paradox of change without difference: A model of change in the arena of fundamental school reform. *Educational Policy, 16*(5), 763-782.

Zint, M. (2002). Comparing three attitude-behavior theories for predicting science teachers' intentions. *Journal of Research in Science Teaching*, *39*(9), 819-844.

Zohar, A., & Nemet, F. (2002). Fostering students' knowledge and argumentation skills through dilemmas in human genetics. *Journal of Research in Science Teaching, 39,* 35–62.

APPENDIX A: TEACHER SURVEY

Dear Participant,

My name is Dr. Meena Balgopal and I am a researcher from Colorado State University in the School of Education. We are conducting a science education research study entitled "Determining the alignment between what teachers are expected to teach, what they know, and how they assess scientific literacy." I am working with my graduate student, Lisa Pitot, and PSD colleagues, Dwayne Schmitz and Dee Dee Wright.

We would like you to take an anonymous online survey. Your participation in this research is voluntary, and will take approximately 30-45 minutes. If you decide to participate in the study, you may withdraw your consent and stop participation at any time without penalty. Once you begin the survey you will have exactly 24 hours to complete it, just make sure you use the same computer to log on again.

We will not collect your name or individual personal identifiers. When we report and share the data to others, we will combine the data from all participants. Principals and other school administrators will not be given the results of any individual teachers and the survey results will not be used in teacher evaluations. We hope to gain more knowledge on the types of professional development needs of PSD secondary science instructors. As an incentive to complete the survey we will be providing all participants that complete the survey a "Stainless Steel Travel Mug with a $5 gift card" donated by the Human Bean (a $20 value). To receive your incentive, you will be redirected to another screen at the end of the survey to provide your name and school location. Your incentive will be delivered to your school following the closing of the survey. Your survey responses will not be connected to your email.

There are no known risks in taking this survey. It is not possible to identify all potential risks in research procedures, but the researcher(s) have taken reasonable safeguards to minimize any known and potential, but unknown, risks.

Your continuation on this survey indicates your consent to participate, and we thank-you in advance for your thoughtful completion of all items.

If you have any questions about the research, please contact:

Dr. Meena Balgopal at (970) 491-4277 or Meena.Balgopal@colostate.edu.

If you have any questions about your rights as a volunteer in this research, contact:

Janell Barker, Human Research Administrator, at 970-491-1655.

TEACHER SURVEY

Part 1. Demographics

a) Which best describes you? () Male () Female

b) What is your (Primary) current teaching assignment?

Choose one () MS 6^{th} () MS 7^{th} () MS 8^{th} () HS 9^{th} () HS 10^{th} () HS 11^{th}-12^{th}

1. How many years have you been teaching? () 1-3; () 4-5; () 6-9; () 10+
2. Choose from the list to indicate your current teaching assignment (District Secondary schools listed)
3. How many years have you been teaching in your current location? () 1-3; () 4-5; () 6-9; () 10+
4. What is your highest level of education? () BS/BA () MS/MA/M.Ed () PhD/EdD

Part II: Scientific Literacy

- How do you know when your students are scientifically literate?

- What are the three most important scientific literacy skills for students to master in a standards-based science course?
- Have you read the Next Generation Science Standards? () Yes; () I have scanned their contents; () I have not read them
- If Yes or "I have scanned their contents" was chosen: In your own words, what are the primary features that differentiate the NGSS's from previous standards (e.g. National Science Education Standards, Colorado Science Education Standards)?

Part IIB: Scientific Literacy in Your Classroom

For the following questions related to scientific literacy skills, please indicate answers based on consideration of the course(s) you teach:

Skill 1: Identify a valid scientific argument (e.g., recognizing when scientific evidence supports a hypothesis).

	Never	1-2 times / semester	1-2 times / quarter	About once a week	Daily
Do you currently teach this skill?	○	○	○	○	○
Do you assess this skill in your class?	○	○	○	○	○

Skill 2: Conduct an effective literature search (e.g., evaluate the validity of sources and distinguish between types of sources)

	Never	1-2 times / semester	1-2 times / quarter	About once a week	Daily
Do you currently teach this skill?	○	○	○	○	○
Do you assess this skill in your class?	○	○	○	○	○

Skill 3: Evaluate the use and misuse of scientific information (e.g., distinguish the appropriate use of science to make societal decisions)

	Never	1-2 times / semester	1-2 times / quarter	About once a week	Daily
Do you currently teach this skill?	○	○	○	○	○
Do you assess this skill in your class?	○	○	○	○	○

Skill 4: Understand elements of research design and how they impact scientific findings/conclusions (e.g., identify strengths and weaknesses in research that are related to bias, sample size, randomization, and experimental control)

	Never	1-2 times / semester	1-2 times / quarter	About once a week	Daily
Do you currently teach this skill?	○	○	○	○	○
Do you assess this skill in your class?	○	○	○	○	○

Skill 5: Create the appropriate graph from data

	Never	1-2 times / semester	1-2 times / quarter	About once a week	Daily
Do you currently teach this skill?	○	○	○	○	○
Do you assess this skill in your class?	○	○	○	○	○

Skill 6: Read and interpret graphical representations of data

	Never	1-2 times / semester	1-2 times / quarter	About once a week	Daily
Do you currently teach this skill?	○	○	○	○	○
Do you assess this skill in your class?	○	○	○	○	○

Skill 7: Solve problems using quantitative skills, including basic statistics (e.g., calculate means, probabilities, percentages, frequencies)

	Never	1-2 times / semester	1-2 times / quarter	About once a week	Daily
Do you currently teach this skill?	○	○	○	○	○
Do you assess this skill in your class?	○	○	○	○	○

Skill 8: Understand and interpret basic statistics (e.g., interpret error bars, understand the need for statistics)

	Never	1-2 times / semester	1-2 times / quarter	About once a week	Daily
Do you currently teach this skill?	○	○	○	○	○
Do you assess this skill in your class?	○	○	○	○	○

Skill 9: Justify inferences, predictions, and conclusions based on quantitative data

	Never	1-2 times / semester	1-2 times / quarter	About once a week	Daily
Do you currently teach this skill?	○	○	○	○	○
Do you assess this skill in your class?	○	○	○	○	○

e). Rank the importance of these same skills to students' scientific literacy development using the likert scale provided**.**

	Not important	Low level of importance	Reasonably Important	Very Important	It is the essential science literacy skill overall the others
Skill 1: Identify a valid scientific argument	○	○	○	○	○
Skill 2: Conduct an effective literature search	○	○	○	○	○
Skill 3: Evaluate the use and misuse of scientific information	○	○	○	○	○
Skill 4: Understand elements of research design and how they impact scientific findings/conclusions	○	○	○	○	○
Skill 5: Create the appropriate graph from data	○	○	○	○	○
Skill 6: Read and interpret graphical representations of data	○	○	○	○	○
Skill 7: Solve problems using quantitative skills, including basic statistics	○	○	○	○	○
Skill 8: Justify inferences, predictions, and conclusions based on quantitative data	○	○	○	○	○
Skill 9: Develop and / or analyze models or modeling	○	○	○	○	○

Part III

The following section contains 28 multiple-choice questions specifically related to the scientific literacy skills listed above. They exemplify how students' scientific literacy skills can be assessed. Please try your best to complete the item*s*, your honest answers will help us better understand teacher's needs related to scientific literacy.

** TOSLS test in Appendix C

Part IV. District Common Assessments

a) Have you participated at all in the development of district common assessments? () Yes; () NO

If yes was chosen the following four questions were presented:

1. Describe your experience and knowledge of developing common* science assessments (*across grade level; within school; within district).
2. Describe any professional development experiences you have had with instrument or assessment development.
3. Why did you choose to participate in the PSD common assessment development process?
4. Describe the process your development team underwent to produce the assessment.

You have now completed the PSD Secondary Science Survey. We thank you very much for your time, continued dedication to science education, and to the students of PSD.

If you would like to receive the incentive offered, please go to the following URL and enter your information on the Google form (You may need to cut and paste the URL).

http://tinyurl.com/science-survey-incentive

APPENDIX B: TOSLS SL SKILLS

Table B1

Correlation of SL skills with TOSLS questions

Table 2. Categories of scientific literacy skills

	Questions	Explanation of skill	Examples of common student challenges and misconceptions
I. Understand methods of inquiry that lead to scientific knowledge			
1. Identify a valid scientific argument	1, 8, 11	Recognize what qualifies as scientific evidence and when scientific evidence supports a hypothesis	Inability to link claims correctly with evidence and lack of scrutiny about evidence "Facts" or even unrelated evidence considered to be support for scientific arguments
2. Evaluate the validity of sources	10, 12, 17, 22, 26	Distinguish between types of sources; identify bias, authority, and reliability	Inability to identify accuracy and credibility issues
3. Evaluate the use and misuse of scientific information	5, 9, 27	Recognize a valid and ethical scientific course of action and identify appropriate use of science by government, industry, and media that is free of bias and economic, and political pressure to make societal decisions	Prevailing political beliefs can dictate how scientific findings are used. All sides of a controversy should be given equal weight regardless of their validity.
4. Understand elements of research design and how they impact scientific findings/conclusions	4, 13, 14	Identify strengths and weaknesses in research design related to bias, sample size, randomization, and experimental control	Misunderstanding randomization contextualized in a particular study design. General lack of understanding of elements of good research design.
II. Organize, analyze, and interpret quantitative data and scientific information			
5. Create graphical representations of data	15	Identify the appropriate format for the graphical representation of data given particular type of data	Scatter plots show differences between groups. Scatter plots are best for representing means, because the graph shows the entire range of data.
6. Read and interpret graphical representations of data	2, 6, 7, 18	Interpret data presented graphically to make a conclusion about study findings	Difficulty in interpreting graphs Inability to match patterns of growth, (e.g., linear or exponential) with graph shape
7. Solve problems using quantitative skills, including probability and statistics	16, 20, 23	Calculate probabilities, percentages, and frequencies to draw a conclusion	Guessing the correct answer without being able to explain basic math calculations Statements indicative of low self-efficacy: "I'm not good at math."
8. Understand and interpret basic statistics	3, 19, 24	Understand the need for statistics to quantify uncertainty in data	Lack of familiarity with function of statistics and with scientific uncertainty. Statistics prove data is correct or true.
9. Justify inferences, predictions, and conclusions based on quantitative data	21, 25, 28	Interpret data and critique experimental designs to evaluate hypotheses and recognize flaws in arguments	Tendency to misinterpret or ignore graphical data when developing a hypothesis or evaluating an argument

APPENDIX C: TEST OF SCIENTIFIC LITERACY SKILLS (TOSLS)

Directions: There are 28 multiple-choice questions. You will have about 35 minutes to work on the questions. Be sure to answer as many of the questions as you can in the time allotted. You will receive attendance points for completing the entire assignment today. Your grade will depend on completeness and thoroughness, not on correct answers. But, try your best, your honest answers will help us better prepare the materials for the remainder of the semester.

Mark your answers on the scantron sheet.

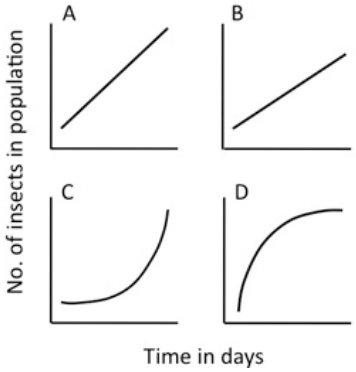
Bubble in your #ID on your scantron.

Do NOT use a calculator. Thank you for your participation in this project!

1. Which of the following is a valid scientific argument?
   a. Measurements of sea level on the Gulf Coast taken this year are lower than normal; the average monthly measurements were almost 0.1 cm lower than normal in some areas. These facts prove that sea level rise is not a problem.
   b. A strain of mice was genetically engineered to lack a certain gene, and the mice were unable to reproduce. Introduction of the gene back into the mutant mice restored their ability to reproduce. These facts indicate that the gene is essential for mouse reproduction.
   c. A poll revealed that 34% of Americans believe that dinosaurs and early humans co-existed because fossil footprints of each species were found in the same location. This widespread belief is appropriate evidence to support the claim that humans did not evolve from ape ancestors.
   d. This winter, the northeastern US received record amounts of snowfall, and the average monthly temperatures were more than 2°F lower than normal in some areas. These facts indicate that climate change is occurring.

2. While growing vegetables in your backyard, you noticed a particular kind of insect eating your plants. You took a rough count (see data below) of the insect population over time. Which graph shows the best representation of your data?

Time (days)	Insect Population (number)
2	7
4	16
8	60
10	123



3. A study about life expectancy was conducted using a random sample of 1,000 participants from the United States. In this sample, the average life expectancy was 80.1 years for females and 74.9 years for males. What is one way that you can **increase your certainty** that women truly live longer than men in the United States' general population?
   a. Subtract the average male life expectancy from the average female expectancy. If the value is positive, females live longer.
   b. Conduct a statistical analysis to determine if females live significantly longer than males.
   c. Graph the mean (average) life expectancy values of females and males and visually analyze the data.
   d. There is no way to increase your certainty that there is a difference between sexes.

4. Which of the following research studies is **least likely** to contain a confounding factor (variable that provides an alternative explanation for results) in its design?
   a. Researchers randomly assign participants to experimental and control groups. Females make up 35% of the experimental group and 75% of the control group.
   b. To explore trends in the spiritual/religious beliefs of students attending U.S. universities, researchers survey a random selection of 500 freshmen at a small private university in the South.
   c. To evaluate the effect of a new diet program, researchers compare weight loss between participants randomly assigned to treatment (diet) and control (no diet) groups, while controlling for average daily exercise and pre-diet weight.
   d. Researchers tested the effectiveness of a new tree fertilizer on 10,000 saplings. Saplings in the control group (no fertilizer) were tested in the fall, whereas the treatment group (fertilizer) were tested the following spring.

5. Which of the following actions is a valid scientific course of action?
   a. A government agency relies heavily on two industry-funded studies in declaring a chemical found in plastics safe for humans, while ignoring studies linking the chemical with adverse health effects.
   b. Journalists give equal credibility to both sides of a scientific story, even though one side has been disproven by many experiments.
   c. A government agency decides to alter public health messages about breast-feeding in response to pressure from a council of businesses involved in manufacturing infant formula.
   d. Several research studies have found a new drug to be effective for treating the symptoms of autism; however, a government agency refuses to approve the drug until long term effects are known.

Background for question 6: The following graph appeared in a scientific article[1] about the effects of pesticides on tadpoles in their natural environment.

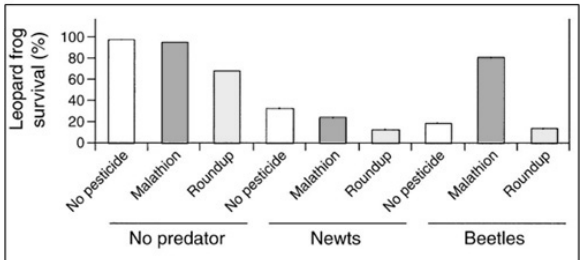
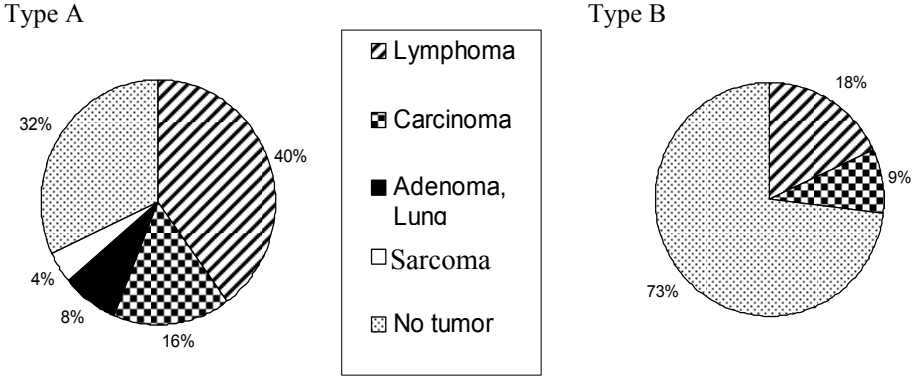


Fig. 2. Total survival of Leopard frog tadpoles living in ponds with 2 pesticides added (Malathion and Roundup) and 2 tadpole predators present (newts and beetles).

6. When beetles were introduced as predators to the Leopard frog tadpoles, and the pesticide Malathion was added, the results were unusual. Which of the following is a plausible hypothesis to explain these results?
   a. The Malathion killed the tadpoles, causing the beetles to be hungrier and eat more tadpoles.
   b. The Malathion killed the tadpoles, so the beetles had more food and their population increased.
   c. The Malathion killed the beetles, causing fewer tadpoles to be eaten.
   d. The Malathion killed the beetles, causing the tadpole population to prey on each other.

[1] Modified from Relyea, R.A., N.M. Schoeppner, J.T. Hoverman. 2005. Pesticides and amphibians: the importance of community context. Ecological Applications 15: 1125-1134

7. Which of the following is the **best** interpretation of the graph below[2]?



Tumors found in type A and type B mice. Pie chart depicts relative incidence of tumors. Numbers outside each slice denote the percentage of specific tumor type.

a. Type "A" mice with Lymphoma were more common than type "A" mice with no tumors.
b. Type "B" mice were more likely to have tumors than type "A" mice.
c. Lymphoma was equally common among type "A" and type "B" mice.
d. Carcinoma was less common than Lymphoma only in type "B" mice.

8. Creators of the Shake Weight, a moving dumbbell, claim that their product can produce "incredible strength!" Which of the additional information below would provide the **strongest evidence** supporting the effectiveness of the Shake Weight for increasing muscle strength?
   a. Survey data indicates that on average, users of the Shake Weight report working out with the product 6 days per week, whereas users of standard dumbbells report working out 3 days per week.
   b. Compared to a resting state, users of the Shake Weight had a 300% increase in blood flow to their muscles when using the product.
   c. Survey data indicates that users of the Shake Weight reported significantly greater muscle tone compared to users of standard dumbbells.
   d. Compared to users of standard dumbbells, users of the Shake Weight were able to lift weights that were significantly heavier at the end of an 8-week trial.

9. Which of the following is **not** an example of an appropriate use of science?
   a. A group of scientists who were asked to review grant proposals based their funding recommendations on the researcher's experience, project plans, and preliminary data from the research proposals submitted.
   b. Scientists are selected to help conduct a government-sponsored research study on global climate change based on their political beliefs.
   c. The Fish & Wildlife Service reviews its list of protected and endangered species in response to new research findings.
   d. The Senate stops funding a widely used sex-education program after studies show limited effectiveness of the program.

[2] Modified from Wang, Y., S. Klumpp, H.M. Amin, H. Liang, J. Li, Z. Estrov, P. Zweidler-McKay, S.J.Brandt, A. Agulnick, L. Nagarajan. 2010. SSBP2 is an *in vivo* tumor suppressor and regulator of LDB1 stability. Oncogene 29: 3044-3053.

Background for question 10: Your interest is piqued by a story about human pheromones on the news. A Google search leads you to the following website:

10. For this website (Eros Foundation), which of the following characteristics is **<u>most important</u>** in your confidence that the resource is accurate or not.
    a. The resource may not be accurate, because appropriate references are not provided.
    b. The resource may not be accurate, because the purpose of the site is to advertise a product.
    c. The resource is likely accurate, because appropriate references are provided.
    d. The resource is likely accurate, because the website's author is reputable.

Background for questions 11 – 14: Use the excerpt below (modified from a recent news report on MSNBC.com) for the next few questions.

"A recent study, following more than 2,500 New Yorkers for 9+ years, found that people who drank diet soda every day had a 61% higher risk of vascular events, including stroke and heart attack, compared to those who avoided diet drinks. For this study, Hannah Gardner's research team randomly surveyed 2,564 New Yorkers about their eating behaviors, exercise habits, as well as cigarette and alcohol consumption. Participants were also given physical check-ups, including blood pressure measurements and blood tests for cholesterol and other factors that might affect the risk for heart attack and stroke. The increased likelihood of vascular events remained even after Gardener and her colleagues accounted for risk factors, such as smoking, high blood pressure and high cholesterol levels. The researchers found no increased risk among people who drank regular soda."

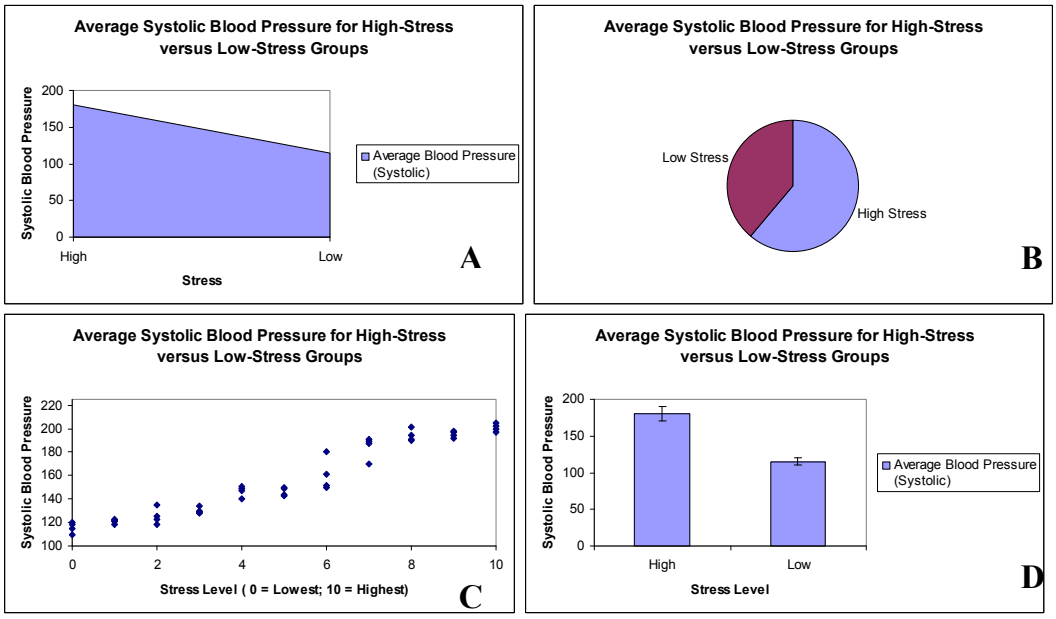
11. The findings of this study suggest that consuming diet soda might lead to increased risk for heart attacks and strokes. From the statements below, identify **additional evidence that supports** this claim:
    a. Findings from an epidemiological study suggest that NYC residents are 6.8 times more likely to die of vascular-related diseases compared to people living in other U.S. cities.
    b. Results from an experimental study demonstrated that individuals randomly assigned to consume one diet soda each day were twice as likely to have a stroke compared to those assigned to drink one regular soda each day.
    c. Animal studies suggest a link between vascular disease and consumption of caramel-containing products (ingredient that gives sodas their dark color).
    d. Survey results indicate that people who drink one or more diet soda each day smoke more frequently than people who drink no diet soda, leading to increases in vascular events.

12. The excerpt above comes from what type of source of information?
    a. Primary (Research studies performed, written and then submitted for peer-review to a scientific journal.)
    b. Secondary (Reviews of several research studies written up as a summary article with references that are submitted to a scientific journal.)
    c. Tertiary (Media reports, encyclopedia entries or documents published by government agencies.)
    d. None of the above

13. The lead researcher was quoted as saying, "I think diet soda drinkers need to stay tuned, but I don't think that anyone should change their behaviors quite yet." Why didn't she warn people to stop drinking diet soda right away?
    a. The results should be replicated with a sample more representative of the U.S. population.
    b. There may be significant confounds present (alternative explanations for the relationship between diet sodas and vascular disease).
    c. Subjects were not randomly assigned to treatment and control groups.
    d. All of the above

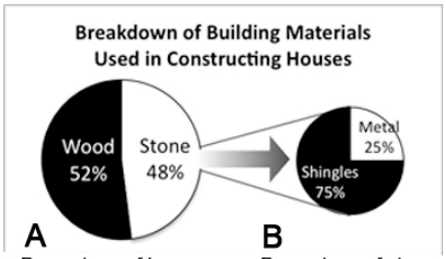
14. Which of the following attributes is **not** a strength of the study's research design?"
    a. Collecting data from a large sample size.
    b. Randomly sampling NYC residents.
    c. Randomly assigning participants to control and experimental groups.
    d. All of the above.

15. Researchers found that chronically stressed individuals have significantly higher blood pressure compared to individuals with little stress. Which graph would be most appropriate for displaying the mean (average) blood pressure scores for high-stress and low-stress groups of people?



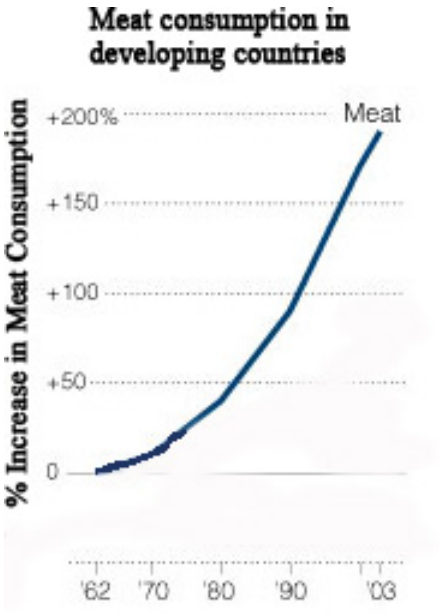
Background for question 16: Energy efficiency of houses depends on the construction materials used and how they are suited to different climates. Data was collected about the types of building materials used in house construction (results shown below). Stone houses are more energy efficient, but to determine if that efficiency depends on roof style, data was also collected on the percentage of stone houses that had either shingles or a metal roof.

16. What proportion of houses were constructed of a stone base with a shingled roof?
    a. 25%
    b. 36%
    c. 48%
    d. Cannot be calculated without knowing the original number of survey participants.



17. The **most important** factor influencing you to categorize a research article as trustworthy science is:
    a. the presence of data or graphs
    b. the article was evaluated by unbiased third-party experts
    c. the reputation of the researchers
    d. the publisher of the article

18. Which of the following is the **most accurate** conclusion you can make from the data in this graph[3]?



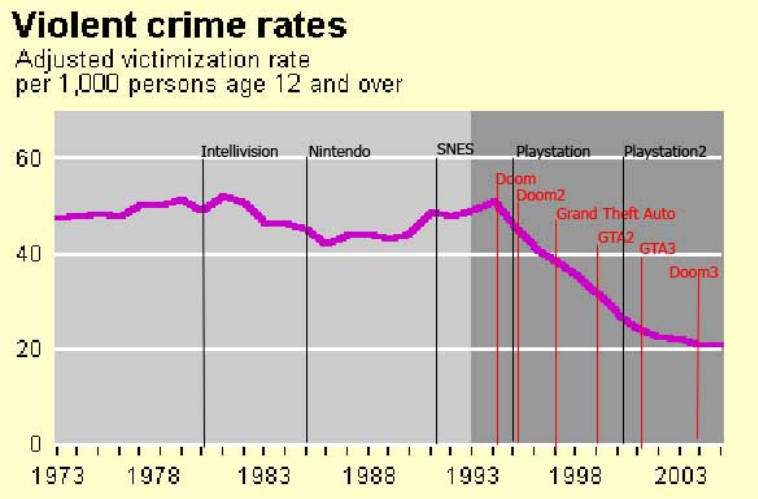
a. The largest increase in meat consumption has occurred in the past 20 years.
b. Meat consumption has increased at a constant rate over the past 40 years.
c. Meat consumption doubles in developing countries every 20 years.
d. Meat consumption increases by 50% every 10 years.

19. Two studies estimate the mean caffeine content of an energy drink. Each study uses the same test on a random sample of the energy drink. Study 1 uses 25 bottles, and study 2 uses 100 bottles. Which statement is true?
    a. The estimate of the actual mean caffeine content from each study will be equally uncertain.
    b. The uncertainty in the estimate of the actual mean caffeine content will be smaller in study 1 than in study 2.
    c. The uncertainty in the estimate of the actual mean caffeine content will be larger in study 1 than in study 2.
    d. None of the above

20. A hurricane wiped out 40% of the wild rats in a coastal city. Then, a disease spread through stagnant water killing 20% of the rats that survived the hurricane. What percentage of the original population of rats is left after these 2 events?
    a. 40%
    b. 48%
    c. 60%
    d. Cannot be calculated without knowing the original number of rats.

[3] Modified from Rosenthal, Elizabeth. 2008. As More Eat Meat, a Bid to Cut Emissions. New York Times, December 3, 2008. Accessed June 9, 2011 http://www.nytimes.com/2008/12/04/science/earth/04meat.html

Background for question 21: A videogame enthusiast argued that playing violent video games (e.g., Doom, Grand Theft Auto) does not cause increases in violent crimes as critics often claim. To support his argument, he presents the graph below. He points out that the rate of violent crimes has decreased dramatically, beginning around the time the first "moderately violent" video game, Doom, was introduced.



21. Considering the information presented in this graph, what is the **most critical flaw** in the blogger's argument?
    a. Violent crime rates appear to increase slightly after the introduction of the Intellivision and SNES game systems.
    b. The graph does not show violent crime rates for children under the age of 12, so results are biased.
    c. The decreasing trend in violent crime rates may be caused by something other than violent video games
    d. The graph only shows data up to 2003. More current data are needed.

22. Your doctor prescribed you a drug that is brand new. The drug has some significant side effects, so you do some research to determine the effectiveness of the new drug compared to similar drugs on the market. Which of the following sources would provide the **most accurate** information?
    a. the drug manufacturer's pamphlet/website
    b. a special feature about the drug on the nightly news
    c. a research study conducted by outside researchers
    d. information from a trusted friend who has been taking the drug for six months

23. A gene test shows promising results in providing early detection for colon cancer. However, 5% of all test results are falsely positive; that is, results indicate that cancer is present when the patient is, in fact, cancer-free. Given this false positive rate, how many people out of 10,000 would have a false positive result and be alarmed unnecessarily?
    a. 5
    b. 35
    c. 50
    d. 500

24. Why do researchers use statistics to draw conclusions about their data?
    a. Researchers usually collect data (information) about everyone/everything in the population.
    b. The public is easily persuaded by numbers and statistics.
    c. The true answers to researchers' questions can only be revealed through statistical analyses.
    d. Researchers are making inferences about a population using estimates from a smaller sample.

25. A researcher hypothesizes that immunizations containing traces of mercury **do not** cause autism in children. Which of the following data provides the **strongest** test of this hypothesis?
    a. a count of the number of children who were immunized and have autism
    b. yearly screening data on autism symptoms for immunized and non-immunized children from birth to age 12
    c. mean (average) rate of autism for children born in the United States
    d. mean (average) blood mercury concentration in children with autism

Background for Question 26: You've been doing research to help your grandmother understand two new drugs for osteoporosis. One publication, *Eurasian Journal of Bone and Joint Medicine*, contains articles with data only showing the effectiveness of one of these new drugs. A pharmaceutical company funded the *Eurasian Journal of Bone and Joint Medicine* production and most advertisements in the journal are for this company's products. In your searches, you find other articles that show the same drug has only limited effectiveness.

26. Pick the **best** answer that would help you decide about the credibility of the *Eurasian Journal of Bone and Joint Medicine*:
    a. It is not a credible source of scientific research because there were advertisements within the journal.
    b. It is a credible source of scientific research because the publication lists reviewers with appropriate credentials who evaluated the quality of the research articles prior to publication.
    c. It is not a credible source of scientific research because only studies showing the effectiveness of the company's drugs were included in the journal.
    d. It is a credible source of scientific research because the studies published in the journal were later replicated by other researchers.

27. Which of the following actions is a valid scientific course of action?
    a. A scientific journal rejects a study because the results provide evidence against a widely accepted model.
    b. The scientific journal, Science, retracts a published article after discovering that the researcher misrepresented the data.
    c. A researcher distributes free samples of a new drug that she is developing to patients in need.
    d. A senior scientist encourages his graduate student to publish a study containing ground-breaking findings that cannot be verified.

Background for question 28: Researchers interested in the relation between River Shrimp (Macrobrachium) abundance and pool site elevation, presented the data in the graph below. Interestingly, the researchers also noted that water pools tended to be shallower at higher elevations.

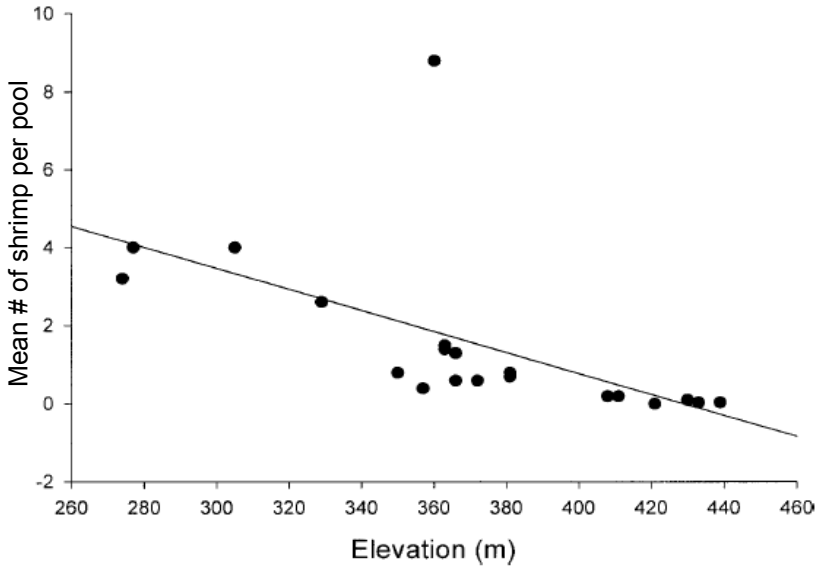


FIG. 3. Relationship between total abundance of *Macrobrachium* (1988–2002) and elevation in Quebrada Prieta.

28. Which of the following is a plausible hypothesis to explain the results presented in the graph?

    a. There are more water pools at elevations above 340 meters because it rains more frequently in higher elevations.
    b. River shrimp are more abundant in lower elevations because pools at these sites tend to be deeper.
    c. This graph cannot be interpreted due to an outlying data point.
    d. As elevation increases, shrimp abundance increases because they have fewer predators at higher elevations.

APPENDIX D: TOSLS SKILLS AND ANSWER KEY

Table D1

TOSLS Answer key

SKILL	DESCRIPTION	Question	Correct Answer
1 (3 Qs)	Identify a valid scientific argument (e.g., recognizing when scientific evidence supports a hypothesis)	1	B
		8	D
		11	B
2 (5 Qs)	Conduct an effective literature search (e.g. Evaluate the validity of sources (e.g., websites, peer reviewed journals) and distinguish between types of sources)	10	B
		12	C
		17	B
		22	C
		26	C
3 (3 Qs)	Evaluate the use and misuse of scientific information (e.g Recognize a valid scientific course of action, distinguish the appropriate use of science to make societal decisions)	5	D
		9	B
		27	B
4 (4 Qs)	Understand elements of research design and how they impact scientific findings/conclusions (e.g. identify strengths and weaknesses in research related to bias, sample size, randomization, experimental control)	4	C
		13	D
		14	C
		25	B
5 (1Q)	Make a graph	15	D
6 (4 Qs)	Read and interpret graphical representations of data	2	C
		6	C
		7	A
		18	A
7 (3 Qs)	Solve problems using quantitative skills, including probability and statistics (e.g calculate means, probabilities, percentages, frequencies)	16	B
		20	B
		23	D
8 (3 Qs)	Understand and interpret basic statistics (e.g. interpret error bars, understand the need for statistics	3	B
		19	C
		24	D
9 (2 Qs)	Justify inferences, predictions, and conclusions based on quantitative data	21	C
		28	B

APPENDIX E: SUMMARY STATISTICS FOR CONTENT ANALYSIS OF STANDARDS

Table E1

Content analysis of CAS

Code	**Colorado HS standards (n=153)**		**Colorado MS standards (n=164)**	
	Frequency	Percent	Frequency	Percent
Claim	145	95	147	90
Evidence	129	84	134	82
Reasoning	54	35	39	24
Source Credibility	48	31	43	26
Society/Tech.	53	35	42	26
Inquiry	38	25	28	17
Data/Graphs	19	12	3	2
Math. Reasoning (Includes basic statistics)	25	16	8	5
Modeling	21	14	33	20
Justification (Numerical)	29	19	10	6

Table E2

Content analysis of NGSS

Code	NGSS HS standards (n=67)		NGSS MS standards (n=60)	
	Frequency	Percent	Frequency	Percent
Claim	67	100	53	88
Evidence	66	99	53	88
Reasoning	29	43	34	57
Source Credibility	10	15	3	5
Society/Tech.	17	25	9	15
Inquiry	11	16	7	12
Data/Graphs	4	6	1	2
Math. Reasoning	20	30	3	5
Modeling	29	43	14	23
Justification (Numerical)	31	46	25	41

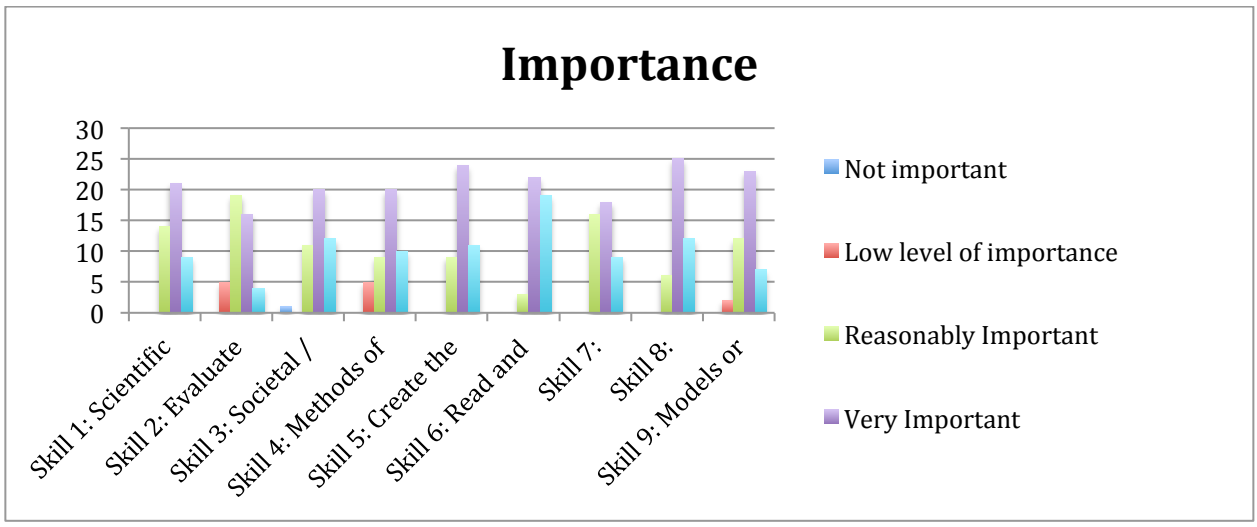
APPENDIX F: ANALYSIS OF TEACHER RANKING OF THE IMPORTANCE OF SL SKILLS AND WHETHER THEY TEACH OR ASSESS THESE SKILLS.

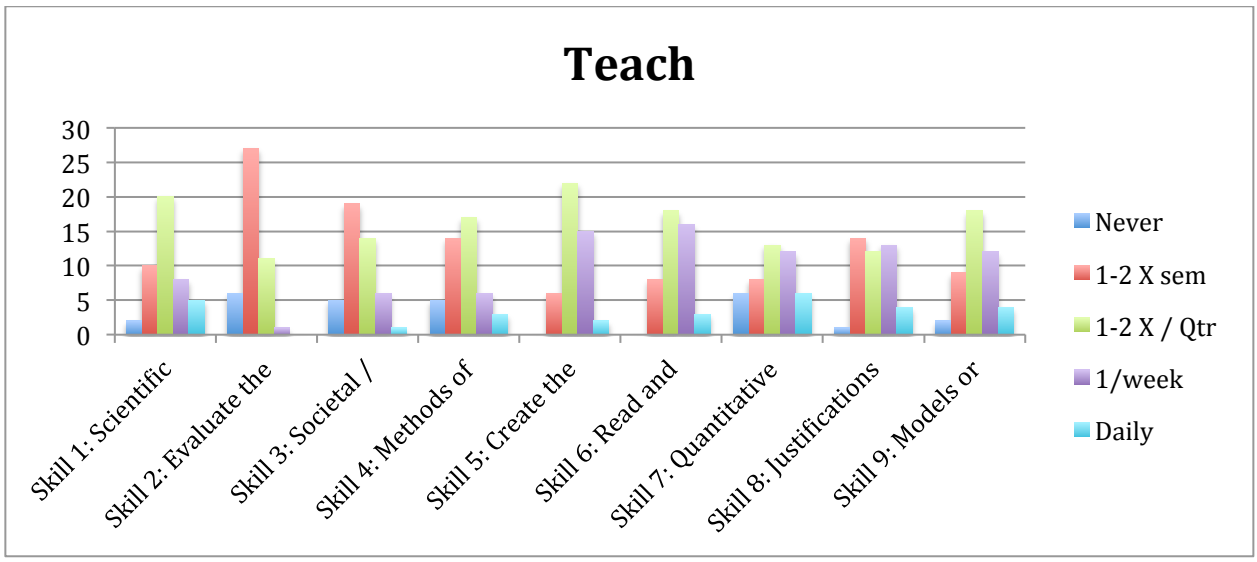
(Spearman Rho values and Graphs of teacher survey data)

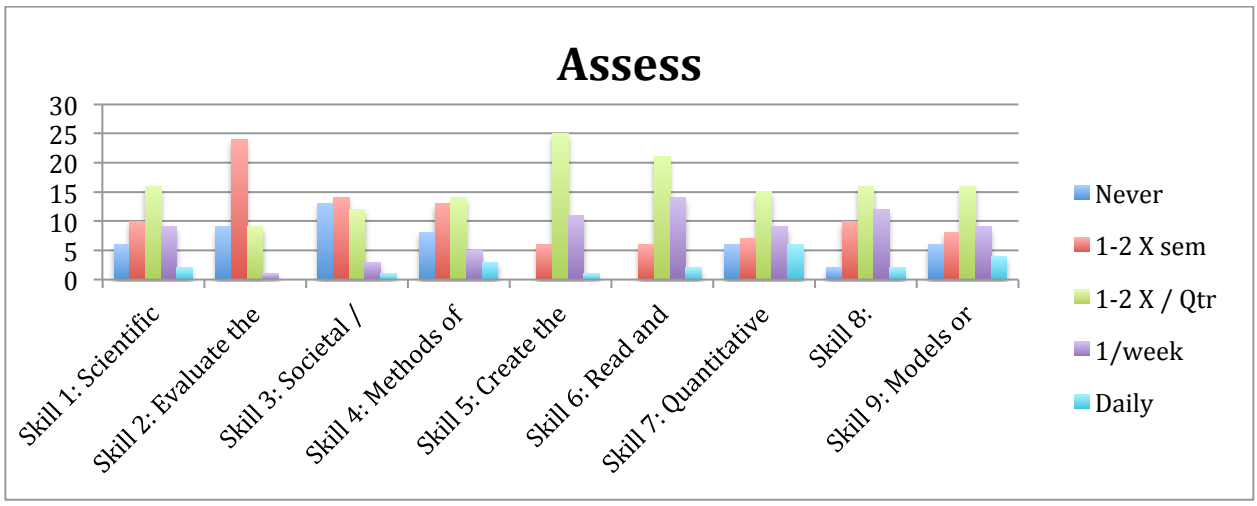
Table F1

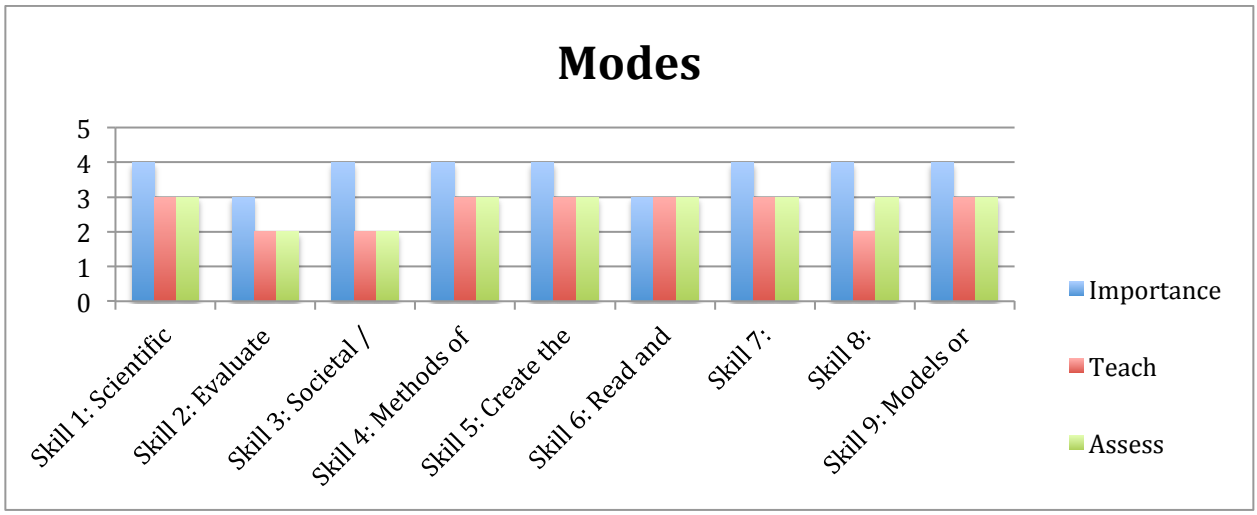
Spearman Rho values

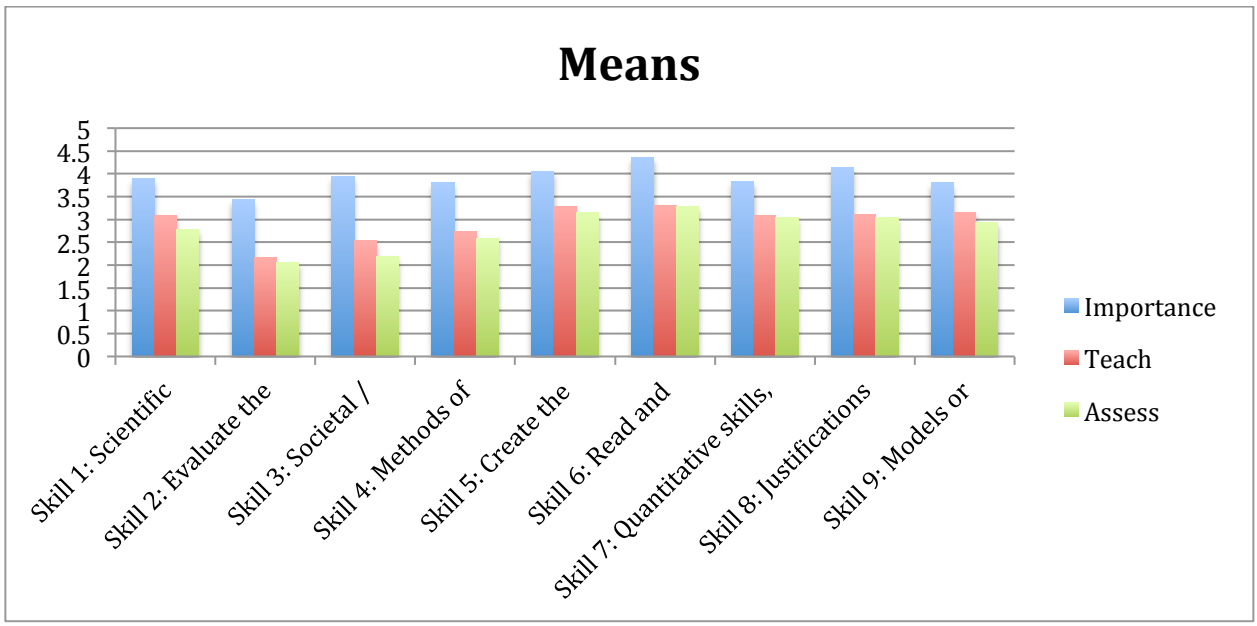
Variable 1: Skill 1-9 Importance vs. Variable 2: Skill 1-9 Teach									
Var 1	Sk1 Imp	Sk2 Imp	Sk3 Imp	Sk4 Imp	Sk5 Imp	Sk6 Imp	Sk7 Imp	Sk8 Imp	Sk9 Imp
Var 2	Sk 1 Teach	Sk2 Teach	Sk3 Teach	Sk4 Teach	Sk5 Teach	Sk6 Teach	Sk7 Teach	Sk8 Teach	Sk9 Teach
r	0.359	0.3	-0.5	-0.1	0.41	0.41	0.658	0	0.8
p-value	0.517	0.683	0.45	0.95	0.45	0.45	0.233	1	0.133
n	5	5	5	5	5	5	5	5	5
Variable 1: Skill 1-9 Importance vs. Variable 2: Skill 1-9 Assess									
Var 1	Sk1 Imp	Sk2 Imp	Sk3 Imp	Sk4 Imp	Sk5 Imp	Sk6 Imp	Sk7 Imp	Sk8 Imp	Sk9 Imp
Var 2	Sk 1 Assess	Sk2 Assess	Sk3 Assess	Sk4 Assess	Sk5 Assess	Sk6 Assess	Sk7 Assess	Sk8 Assess	Sk9 Assess
r	0.205	-0.1	**-0.9**	-0.395	0.41	0.41	0.564	0.289	0.6
p-value	0.683	0.95	**0.083**	0.517	0.45	0.45	0.35	0.683	0.35
n	5	5	5	5	5	5	5	5	5


Importance
30
25
20
15
10
5
0
Skill 1: Scientific
Skill 2: Evaluate
Skill 3: Societal /
Skill 4: Methods of
Skill 5: Create the
Skill 6: Read and
Skill 7:
Skill 8:
Skill 9: Models or
Not important
Low level of importance
Reasonably Important
Very Important


Teach
30
25
20
15
10
5
0
Skill 1: Scientific
Skill 2: Evaluate the
Skill 3: Societal /
Skill 4: Methods of
Skill 5: Create the
Skill 6: Read and
Skill 7: Quantitative
Skill 8: Justifications
Skill 9: Models or
Never
1-2 X sem
1-2 X / Qtr
1/week
Daily


Assess
30
25
20
15
10
5
0
Skill 1: Scientific
Skill 2: Evaluate the
Skill 3: Societal /
Skill 4: Methods of
Skill 5: Create the
Skill 6: Read and
Skill 7: Quantitative
Skill 8:
Skill 9: Models or
Never
1-2 X sem
1-2 X / Qtr
1/week
Daily


Modes
5
4
3
2
1
0
Skill 1: Scientific
Skill 2: Evaluate
Skill 3: Societal /
Skill 4: Methods of
Skill 5: Create the
Skill 6: Read and
Skill 7:
Skill 8:
Skill 9: Models or
Importance
Teach
Assess


Means
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
Skill 1: Scientific
Skill 2: Evaluate the
Skill 3: Societal /
Skill 4: Methods of
Skill 5: Create the
Skill 6: Read and
Skill 7: Quantitative skills,
Skill 8: Justifications
Skill 9: Models or
Importance
Teach
Assess

APPENDIX G: COMMON ASSESSMENT ALIGNMENT DATA

Table G1

6th grade common assessment PI data

6th Grade									
A priori code	Physical Science Exam			Life Science – Ecology Exam			Earth Science Exam		
	Exam	CAS	PI	Exam	CAS	PI	Exam	CAS	PI
1.	0	.29	.86	.02	.23	.90	0	.22	.89
1a.	.86	.21	.68	.82	.32	.75	.82	.18	.68
2.	0	.05	.98	0	.04	.98	0	.16	.92
3.	0	.06	.97	.022	.24	.89	.093	.25	.92
4.	.01	.13	.94	0	.04	.98	0	.06	.97
5/6	.03	.00	.98	.122	0	.94	.074	0	.96
7/8	.03	.10	.97	0	0	1	.017	.04	.99
9.	0	0	1	0	0	1	0	.03	.99
10.	.06	.15	.96	.017	.13	.94	0	.06	.97
	Overall PI =		0.32	Overall PI =		.38	Overall PI =		.29

Table G2

7th grade common assessment PI data

7th Grade						
A priori code	Cells & Genetics Exams			Earth Science – Geology Exam		
	Exam	CAS	PI	Exam	CAS	PI
1.	.09	.23	.93	.13	.27	.93
1a.	.723	.33	.80	.623	.36	.87
2.	0	.10	.95	0	.03	.985
3.	.019	.10	.96	.0122	.06	.976
4.	.033	.10	.97	.111	.10	.995
5/6	.0966	0	.95	.068	.04	.986
7/8	.0089	.06	.97	.0556	.04	.992
9.	0	.01	1.0	0	.03	.985
10.	.027	.06	.98	0	.08	.96
	Overall PI =		.51	Overall PI =		.50

Table G3

8th grade common assessment PI values

8th Grade									
A priori code	Physical Science Exams: Motion and Forces; Energy and Waves			Earth Science – Weather and Climate Exam			Earth Science Astronomy Exam		
	Exam	CAS	PI	Exam	CAS	PI	Exam	CAS	PI
1.	0	.19	.91	0	.13	.94	0	.23	.89
1a.	.83	.33	.75	.85	.31	.73	.80	.26	.73
2.	0	.10	.95	0	.12	.94	0	.16	.92
3.	0	.04	.98	.007	.26	.87	.015	.15	.93
4.	.015	.13	.94	0	.03	.985	.03	.04	.995
5/6	.045	.05	1.0	.11	.06	.975	.142	.03	.94
7/8	.03	.10	.97	.008	.03	.99	0	0	1
9.	.067	.056	.99	0	0	1	0	0	1
10.	0	.03	.99	.034	.05	.99	.01	.20	.905
	Overall PI = 0.47			Overall PI =		.42	Overall PI =		.34

Table G4

High School common assessment PI values

High School Common Assessment Alignment Data									
A priori code	Biology			Earth Systems Science			Chemistry		
	Exam	CAS	PI	Exam	CAS	PI	Exam	CAS	PI
1.	O	.287	.86	0	.28	.86	0	.25	.86
1a.	.79	.16	.69	.78	.17	.70	.93	.22	.70
2.	0	.112	.94	0	.091	.95	0	.061	.97
3.	.04	.136	.95	.046	.121	.96	0	.048	.98
4.	.033	.091	.97	.015	.098	.96	0	.122	.94
5/6	.102	.025	.96	.103	.049	.97	.0083	.042	.98
7/8	.017	.063	.98	.059	.098	.98	.167	.12	.98
9.	0	.058	.97	0	0	1	0	0	1
10.	.021	.072	.97	0	.04	.98	0	.049	.98
	Overall PI		.29	Overall PI		.32	Overall PI		.39