PREDICTING SCIENCE LITERACY: A MULTIPLE REGRESSION MODEL OF FACTORS THAT INFLUENCE SCIENCE LITERACY

A Dissertation

by

JEFFREY R. CHANDLER

Submitted to the Graduate School
of Texas A&M University-Commerce
in partial fulfillment of the requirements
for the degree of
DOCTOR OF EDUCATION
August 2020

PREDICTING SCIENCE LITERACY: A MULTIPLE REGRESSION MODEL OF FACTORS THAT INFLUENCE SCIENCE LITERACY

A Dissertation

by

JEFFREY R. CHANDLER

Approved by:

Advisor: Gilbert Naizer

Committee: David Brown
Becky Sinclair
Stephen Starnes

Head of Department: Jennifer Dyer

Dean of the College: Kimberly McLeod

Dean of the Graduate School: Jennifer L. Schroeder

ABSTRACT

PREDICTING SCIENCE LITERACY: A MULTIPLE REGRESSION MODEL OF FACTORS THAT INFLUENCE SCIENCE LITERACY

Jeffrey R. Chandler, EdD
Texas A&M University-Commerce, 2020

Advisor: Gilbert Naizer, PhD

Academic scholarship has focused on the development of an operational definition (Bybee, 1997; Laugksch, 2000; Miller, 1998), classification system (Roberts, 2007), or intervention (Allchin, Andersen, & Nielsen, 2014) of science literacy. The present study investigated the relationship between science literacy, as measured by the Test of Science Literacy Skills, and predictive factors commonly collected in secondary schools. This quantitative correlation study examined the relationship between variables commonly collected by schools with science literacy performance to quantify the impact of and target intervention at variables most likely to impact science literacy. Predictive factors included Gender, Ethnicity, Economic Need, English Proficiency, Number of Science Courses Completed, Biology Content Knowledge, English Language Arts Content Knowledge, Algebra Content Knowledge, and Attitude Toward Science. Participants were selected through convenience sampling from a single suburban high school in the southwestern United States. A hierarchal, multiple regression model was developed using the Theory of Reasoned Action and Planned Behavior framework (Fishbein & Ajzen, 2010). The model predicted science literacy based on distal, opportunity, propensity, and demographic factors. A significant regression equation was found. Results suggest that students' science literacy scores can increase by increasing the number of science courses complete, facilitating interventions via English Language Arts, and supporting students' development of a positive attitude toward science.

ACKNOWLEDGEMENTS

This dissertation would not have been possible without the support of my family, friends, colleagues, and committee. I would like to thank my parents for instilling the importance of education, the expectation of success, and a supportive environment to grow in. I would like to thank my children for understanding and supporting the continuation of my education. My wife provided the initial motivation and was critical to the completion of this dissertation. She worked tirelessly to support our family managing our children, supporting the household and as a sounding board for experimental design and statistical analysis of this project.

I would like to thank the friends and colleagues for providing an environment that supports continued education. They provided a great deal of support in the data acquisition phase and providing feedback. I am fortunate in having the support of two friends and colleagues that challenged me through thought provoking discussion, debate, and a love of science that made me a better teacher and research. Thank you, Leonard and Wolowitz.

This dissertation would not have been possible without the support of the dissertation committee. Dr. David Brown for sparking my interest in literacy and science literacy as well as differentiating coursework to support my continued growth. Dr. Stephen Starnes for supporting my growth in science content knowledge of linking academic content with science literacy. Dr. Becky Sinclair for refining my experimental protocol and the encouragement to avoid limitations to the study. Finally, I would like to thank my dissertation committee chair Dr. Gilbert Naizer for the countless hours he has spent in support of this dissertation. Dr. Naizer guided my scattered thoughts, provided key ideas, while allowing me to explore and make this study my own. He knew exactly when to prod and when to let me discover on my own. Thank you all for making this possible.

TABLE OF CONTENTS

CHAPTER

CHAPTER

LIST OF TABLES

TABLE

TABLE

LIST OF FIGURES

FIGURE

Chapter 1

INTRODUCTION

The development of nuclear power at the conclusion of World War II secured the scientific hegemony of the United States in the immediate aftermath. On October 4, 1957, the perception of American scientific superiority came into question with the Soviet Union launch of Sputnik. The relegation of the United States to a secondary position in scientific pursuits led to renewed interest in scientific education. During this time, Paul Hurd coined the term *science literacy* in stating "we are raising a new generation of Americans that is scientifically and technologically illiterate" (National Commission of Excellence in Education, 1983, p. 12).
In 1958, The National Defense Education Act was passed, which declared that the current state of education was insufficient and "national interests require . . . that the Federal Government give assistance to education for programs, which are important to our defense" (National Defense Education Act of 1958, §101). From 1958 to 1973, there was a 5% increase in the number of students enrolled in science courses for grades 7-12 (Helgeson, Blosser, & Howe, 1977). A large influx of financial support from the Federal government resulted in the development of a science curriculum and resources which in turn sparked major advances in telecommunications and electronics, culminating with a successful landing of an astronaut on the moon in 1969.

During the 1970s, a downward trend in math and science scores led to a national commission examining the state of education. A report by the National Commission of Excellence in Education entitled *A Nation at Risk* indicated that American students' performance ranked the lowest among industrial nations in seven key international assessments. Thirteen percent of Americans were functionally illiterate; SAT scores had dropped by 40 points in math and 50 points in reading, and there was a steady decline in science achievement scores among high school seniors (National Commission of Excellence in Education, 1983). US students' scores in scientific literacy, as measured by the Trends in Mathematics and Science Survey (TIMSS) and the Program for International Student Assessment (PISA), consistently remained among the lowest of economically developed countries. In 1985, as a direct response to the findings of *A Nation at Risk*, the American Association for the Advancement of Science (AAAS) began Project 2061 with the goal of enhancing science education by promoting literacy in science, math, and technology (AAAS, 1990). The reforms in the 1980's halted the decline in science scores while providing a slight increase in advanced performance (Campbell, Voelkl, & Donahue, 1997).

Performance slightly improved in the early nineties but still fell significantly below previous scores (Campbell et al., 1997). Key ideas from Project 2061 were implemented into suggested learning standards outlined by the National Research Council (NRC, 1996). Science education focused on "achieving the goal of scientific literacy for all students that is described in the content standards" (NRC, 1996, p. 1). The suggestions of the NRC were implemented inconsistently in state learning standards and failed to result in appreciable gains with NAEP scores falling in the 2000 administration of NAEP Science (Campbell, Hombo, & Mazzeo, 2000).

No Child Left Behind (NCLB; 2002) legislation, passed in 2001, attempted to address performance gaps by increasing the accountability of schools with yearly assessments, higher standards for teachers, and federal guidelines of student performance. Despite a 29% increase in federal funding for education, the 2005 and 2009 NAEP Science scores stagnated (NCES, 2011). This led to the current iteration of the science curriculum, the Next Generation Science Standards

(NGSS), which is a comprehensive K-12 science curriculum emphasizing science literacy (NRC, 2013). Forty states have adopted NGSS, which has resulted in a quasi-national science curriculum. The goals of the NGSS heavily focus on the development and application of science literacy.

The emphasis on science literacy in curriculum, national, and international assessments has led to questions about the operational definition of science literacy, core components, and measurement and factors that influence it. Science literacy is a highly complex, multimodal construct that requires the integration of skills from multiple cognitive domains (Allchin, Andersen, & Nielsen, 2014; Tang, 2015). Science literacy is developed not only by the acquisition of relevant skills, but also specific attitudes, normative beliefs, and achievement variables (Summers & Abd-El-Khalick, 2018).

International performance of high school students has remained stagnant despite numerous policy and funding shifts. Students' failure to make tangible gains in science literacy has led to numerous pieces of legislation that have provided additional funds to expand federal support of science education. The current versions of the Elementary and Secondary Education Act (ESEA) and Every Student Succeeds Act (ESSA) require that states administer content-specific standardized tests, including a high school science assessment. Addressing science literacy through a skill-based curriculum intervention has resulted in varied educational outcomes (Faulkner, 2012; Roblin, Schunn, & McKenney, 2018) that are mitigated over time (Impey, Formanek, Sanlyn & Wenger, 2017, NAEP 2015b). Efforts to influence science literacy at the elementary (Shaw, Lyon, Stoddart, Mosqueda, & Menon, 2014; Tong et al., 2014) and undergraduate levels (Graff, 2016) have demonstrated minimal results.

Before effective intervention can be planned, it is necessary to identify and quantify the factors that influence science literacy at the secondary level. The development of a model that incorporates affective, achievement, and demographic factors to predict science literacy can facilitate the development of targeted educational interventions.

Statement of the Problem

The concept of science literacy has been discussed in the literature for over 60 years without a consensus regarding an operational definition of the term. As Feinstein (2011) has stated, the "bulk of the scholarship in this domain consists of speculative descriptions, more concerned with painting a compelling picture than justifying or articulating that picture" (p. 170). Science literacy can be divided into several classifications including civil science literacy, cultural literacy, and functional science literacy. Definitions of science literacy represent an umbrella term for a set of operational subskills coalescing into a single unitary factor (Gormally, Brickman, & Lutz, 2012). Definitions shift in emphasis on operational subskills between content knowledge, process skills, and level of expertise. The incorporation of a nebulous term in the curriculum has resulted in questions about the efficacy of science literacy as a goal (Feinstein, 2011), which seems to serve as a mere slogan for science education (Roberts, 2007). Currently, there is no consensus regarding the definition of science literacy; however, components such as the nature of science, process skills, critical reading, the scientific method, drawing conclusions, and interpreting data independent of content knowledge are generally accepted as indicators of science literacy (Feinstein, 2011; Gormally et al., 2012; Nuhfer et al., 2016).

Science education researchers continue to measure science literacy despite the lack of a universal operational definition. A number of assessments have been developed to quantify scientific literacy (Gormally et al., 2012; Nuhfer et al., 2016; Wenning, 2006) with each test emphasizing a slightly different subset of skills based on students' age and operational definition. Researchers, teachers, and curriculum developers have developed interventions to improve science literacy, but the lack of understanding of factors that influence literacy has limited progress. These factors may impact the development of science literacy (Impey et al., 2017); however, previous research conflates scientific achievement or content area literacy with science literacy (Moss, 2005). Effective intervention should begin by identifying and quantifying factors that influence science literacy (Hattie, Biggs, & Purdie, 1996).

Purpose of the Study

In an effort to improve US students' scores on international assessments, the Department of Education has increased funding for science education, provided curriculum support from the NRC and NGSS, required state accountability for science performance, and required states to increase the number of courses students complete to graduate. The national and international focus on science literacy has led to its incorporation into state and national standards (Kena et al., 2016; NRC, 2013; Organisation for Economic Co-Operation and Development [OECD], 2016). Despite defining science literacy as the goal of science education, a 27-year longitudinal study of 12,600 University of Arizona undergraduate students showed no significant changes in the science literacy of US students (Impey et al., 2017).

Literature on science literacy has focused on establishing an operational definition of science literacy, incorporating it into K-12 curriculum, assessing literacy, or developing interventions to improve literacy. There is a gap in the literature regarding the factors that influence science literacy. The complex, multimodal nature of science literacy suggests that it may be influenced by distal, opportunity, propensity and demographic factors. The development of a model that quantifies the impact of factors that coalesce in science literacy will support the development of effective intervention programs.

The Test of Science Literacy Skills (TOSLS) was originally developed to assess the science literacy of undergraduate students enrolled in an introductory biology course (Gormally, et al., 2012), but has been used in high school settings as well (Chandler, 2017). The TOSLS is designed to measure science literacy independent of content knowledge. The development of a model in this study includes information routinely collected by schools that previous research has suggested influence science literacy or science achievement.

Research Questions

The research questions below guided the development of a predictive model of science literacy as measured by the TOSLS.

How much will variation in demographic, distal, propensity, and opportunity factors predict science literacy as measured by the TOSLS?

1. What is the relationship between distal factors (Economic Need, English Proficiency, and Grade Point Average) and science literacy?
2. What is the relationship between opportunity factors (Number of Science Courses Completed) and science literacy?
3. What is the relationship between propensity factors (Attitude Toward Science and multi-subject content knowledge) and science literacy?
4. What is the relationship between demographic factors (ethnicity and gender) and science literacy?

This study includes factors classified as content knowledge, demographic, and affective variables based on a review of literatures. These variables were reclassified to distal, opportunity, propensity, and demographic variables consistent with the Theory of Reasoned Action Planned Behavior model for predicting behavior (Ajzen, 2012). The classification of factors using this model allows for the hierarchical inclusion in the model consistent with chronological impact on an individual (Byrnes & Miller, 2007) as shown in Table 1.

Table 1

Classification of Variables in Literature and TRAPB Model

Variable	Literature	TRAPB Model
Attitude Toward Science	Affective	Propensity
Economic Need	Demographic	Distal
English Proficiency	Demographic	Distal
Ethnicity	Demographic	Demographic
Gender	Demographic	Demographic
Grade Point Average	Affective	Distal
Science Courses Completed	Affective	Opportunity
STAAR Algebra 1	Content Knowledge	Propensity
STAAR Biology	Content Knowledge	Propensity
STAAR ELA 1	Content Knowledge	Propensity

Hypothesis

The following hypothesis was generated based on the TRAPB theoretical framework to group variables.

$H_{1,0}$ – Distal, opportunity, propensity and demographic factors will not result in statistically significance variance in science literacy as measured by the TOSLS.

Significance of Study

High school science performance has declined from peaks fifty years ago, briefly rebounding before stagnating (NAEP, 2015a; OECD, 2019). Policymakers have attempted to boost poor performance with increased federal and state funding, additional oversight of science programs (National Defense Education Act of 1958, 2018; NCLB, 2002, 2018), adoption of new curricula with emphasis on scientific inquiry (AAAS, 1990, NRC, 2013), and changing graduation requirements to include more science classes (Plunk, Tate, Bierut, & Grucza, 2014). Despite these interventions, performance has remained stagnant with limited growth in student proficiency from 1986 to today (NAEP, 2015a).

An increase in science literacy has the potential to reduce the shortage of highly qualified science and engineering workers needed to sustain the U.S. economy (National Academy of Science, 2007). New technologies have replaced industrialization as key economic indicators that improve the standard of living (NCMST, 2000). The emphasis on transferable skills in science literacy supports the rapid change-of-pace demanded by current economic conditions.

Recently, there has been a resurgence in interest in science literacy with a substantial increase in the number of academic articles published on the topic compared to ten years ago. Educational literature about science literacy has largely focused on definitions (AAAS, 1990; Feinstein, 2011; Miller, 1998), measurement of the construct (Germann, 1988; Gormally et al., 2012; Nuhfer et al., 2016; OECD, 2016), or designing interventions that impact science literacy scores (Auerbach & Schussler, 2017; Foster & Shiel-Rolle, 2011; Guzzetti & Bang, 2010). Interventions have been developed with additional exposure to scientific processes in multiple settings (Newell, Zientek, Tharp, Vogt, & Moreno, 2015). This study will examine the factors most likely to produce long-term increase in students' science literacy.

The results of this study could benefit curriculum developers by clarifying the interdisciplinary relationship between English Language Arts, Algebra, and Science Attitude with science literacy. Science has historically related to mathematics (Wang, 2005), but the impact of reading and writing has been shown to be critical in the development of science literacy for elementary students (Shaw et al., 2014). This study includes cognitive, affective and demographic factors reclassified as distal, opportunity, propensity, and demographic factors.

District leadership and policymakers could benefit from a better understanding of factors that impact science literacy because schools and districts receive funding based on overall student success and narrowing achievement gaps (ESSA, 2015). Targeted intervention prior to the manifestation of demographic variation may narrow existing achievement gaps. Higher performance on state and national assessments related to science literacy could result in greater funding.

Students could also benefit from a better understanding of the factors that influence science literacy. State and national standards have incorporated science literacy as the goal for science education, but curriculum and lessons have not matched those expectations. Student performance on national and international assessments may improve with a more targeted approach. Higher levels of science literacy have been linked to greater science participation, which can lead to greater success (George, 2006). The development of a multiple regression statistical model that incorporates numerous factors accessible to and impacted by educational institutions has the potential to support increased levels of science literacy for high school students.

Method of Procedure

This correlational study includes data collected from students' educational records, scores on Attitude Toward Science, and Science Literacy to generate a hierarchal multiple regression model. The multiple regression model was developed based on factors demonstrated to show significant difference in science literacy or science achievement. Demographic variables collected include Gender, Ethnicity, English Proficiency, and Economic Need. Knowledge variables include Number of Science Courses Completed, STAAR ELA 1, STAAR Algebra 1, and STAAR Biology content scores. Affective variables include Grade Point Average (GPA) and Attitude Toward Science. This approach allowed for a deeper understanding of science literacy by quantifying the factors that impact it.

Selection of Sample

The participant pool was a sample of convenience consisting of 1279 students enrolled in grades 10-12 in a suburban school district outside a major metropolitan area in Texas. The district was selected due to a high state rating, A-level, yet declining post-secondary completion rates. Students enrolling in a college or university increased from 67% in 2005 to 75% in 2007, while the number of students that remain after two years fell from 84% to 78% (National Student Clearinghouse, 2015). The ACT profile report demonstrates students struggle with college-level science work; only 56% of students meet the standard set by the ACT as ready for college-level coursework (ACT, 2015).

The participant population was a stratified sample with the limitation of intact classes. Participants were excluded from the study if they lacked one or more of the factors included in the model. This limited the available participants to sophomores, juniors, and seniors enrolled in a non-biology science course during the 2019-2020 school year. A total of 1279 students were eligible to participate in the study with 445 returning the required assent and consent forms. Sixty-nine students were excluded from the data set for missing information resulting in a total of 376 participants.

Collection of Data

During the final two weeks of the Fall Semester of 2019, teacher permission (n = 12) was obtained to conduct the study during class. The researcher explained the purpose of the study to students and distributed the student assent and parental consent forms during the first week of the Spring Semester. Students were informed that participation in the study would be conducted in class but that it would not impact their course grade. Five students, randomly selected from those returning forms, received a gift card of $20 as an incentive to participate in the study. Students were provided two weeks to return consent and assent forms.

At the end of the two-week window, data collection for the TOSLS and Behavior, Related Attitudes, and Intentions Towards Science (BRAINS) assessment was conducted during normal class hours. The assessments were conducted on Google Classroom with the BRAINS assessment given on Day 1 followed by the TOSLS on Day 2. The modified block schedule adopted by the site resulted in a 24-hour delay in the administration of the TOSLS for some classes. This is a potential confounding variable as student attitudes toward science can be influenced by the school schedule (Spellman & Oliver, 2001). The BRAINS (see Appendix A) and TOSLS (see Appendix B) were digitized and converted into a Google Form for ease of distribution and data collection. The assessments were collected into a single spreadsheet and automatically scored authors, including reverse scoring (Gormally et al., 2012; Summers & Abd-El-Khalick, 2018).

Data aggregation began with the removal of data from students that opted out, failed to return student assent forms, or failed to return parental consent forms. A review of student records was conducted for the remaining participants. The Skyward Student Information System was used to obtain information about individual participants' Gender, Ethnicity, Economic Need, English Proficiency, GPA, and Number of Science Courses Completed. STAAR scores were obtained for English, Algebra 1, and Biology from the local database or state database for students (n = 36) missing STAAR scores for any subject. A review of records was conducted in lieu of direct questioning as a more accurate means of obtaining information consistent with the GPA (Kuncel, Credé, & Thomas, 2005).

Treatment of Data

The researcher collected variables into a single Excel spreadsheet and performed a fidelity check on the sample. A sample (n = 45) of the data collected was reviewed by the researcher from primary sources to ensure the accuracy of the information. Information from participants was removed when information necessary for inclusion in the model was missing. Three errors were found: an incorrect TOSLS score, a missing CTE credit, and a transposition of numbers indicating an error rate of less than 1%. Identifiable information as outlined in the Institutional Review Board protocol was removed before the information was imported to SPSS 26 for analysis. The remaining data was analyzed for outliers with all data falling within the established interquartile range. A Kolmogorov-Smirnov test was conducted to ensure the homogeneity of variance in TOSLS scores (p = .000). Factors were independently assessed against science literacy for statistical significance using t-Test, ANOVA, or Pearson regression analysis respectively. Factors were added in a hierarchal multiple regression based on the distal, opportunity, propensity model outlined by Byrnes and Miller (2007).

Definitions of Terms

The following terms and operational definitions relate to the present investigation.

Attitude Toward Science. Attitude toward science is defined as a unitary construct consisting of a "learned disposition to respond in a consistently favorable or unfavorable manner" toward science (Fishbein & Ajzen, 1975, p. 193).

Content Knowledge. Content knowledge refers to "the major concepts, key themes, multiple perspectives, assumptions, processes of inquiry, structure, and real-world applications of [students'] grade-level and subject-area content" (Texas Administrative Code, 2014).

Economic Need. Students' state of economic need was determined by their participation in the free and reduced school lunch program due to the "strong average relation between the FRPL data and the Census 2000 poverty data" (Cruse & Powers, 2004, p. 6).

Limited English Proficiency. Students identified with Limited English Proficiency "means a student whose primary language is other than English and whose English language skills are such that the student has difficulty performing ordinary classwork in English" (Texas Education Code, 1995).

Science Literacy. According to the Program for International Student Assessment, *scientific literacy* is "the ability to engage with science-related issues, and with the ideas of science, as a reflective citizen" (OECD, 2017, p. 1). Students who are literate in science can "explain phenomena scientifically, evaluate and design scientific enquiry, and interpret data and evidence" (OECD, 2017, p. 1).

Theory of Reasoned Action and Planned Behavior (***TRAPB).*** TRAPB is a theoretical framework that builds on expectancy-value theory in which "intention to perform a given behavior is based on a particular combination of attitudinal, normative, and control considerations" (Fishbein & Ajzen, 2010, p. 22).

Limitations

One of the limitations of this study is that the participant population of this study is substantially different from the general, US student population. The student-participants were selected from a single secondary school, in a suburban district outside a major metropolitan area in Texas with a total enrollment of approximately 2700 students (TAPR, 2018). For the year of 2016, the year for which the most recent demographic data is available, the campus differed from the rest of the state in its percentage of minority, economically disadvantaged, and English language learner students when compared to the national average (see Table 2).

Table 2

Comparison of Student Population in Science High School With State Averages

Group	School	State Average
African American	6.9%	12.6%
Hispanic	20.8%	52.2%
Asian	3.1%	4.0%
Economically disadvantaged	24.6%	59.0%
English language learners	7.3%	18.5%

Note. Data derived from Texas Academic Performance Report. (2018, December). Retrieved from https://rptsvr1.tea.texas.gov/perfreport/tapr/2018/srch.html?srch=C

In addition to ethnicity, economically disadvantaged students as well as those who are English language learners are also underrepresented (TAPR, 2018). The discrepancy in demographic characteristics limits the generalizability of the findings to the rest of the state. Another potential limitation of this study is the impact course enrollments and schedules have on student attitudes toward science. As J. Osborne, Simon and Collins (2003) have explained, attitude toward science varies between courses as students struggle to relate to information presented. The modified block scheduled, with 45-minute everyday classes or 90-minute alternating classes, impacts students' time on task depending on the course schedule. Time spent on task has been shown to impact student attitudes toward science (Spellman & Oliver, 2001).

The results of the BRAINS and TOSLS assessments depend on students answering all questions truthfully and to the best of their abilities. The lack of a grade associated with the assignment may produce lower levels of effort; however, Segarra et al. (2018) have demonstrated that performance on the TOSLS has little variation despite the degree to which the assignment counts toward students' final grade (Segarra et al., 2018).

Delimitations

This study uses a limited, sample of convenience in which the researcher has an established relationship with teachers. The sample is limited to high school students between 10^{th} and 12^{th} grades due to the year in which variables of interest are collected. Information was collected from a single site with demographics that vary from state and national averages. The limited number of participants in key demographic categories related to ethnicity and English language proficiency could impact the overall model. The model included participants who have completed Biology, English 1, and Algebra 1 STAAR assessments excluding freshmen, sophomores, private school transfers, and out-of-state transfers.

Assumptions

This study assumes the following:

1. Participants answered survey questions truthfully and to the best of their abilities.
2. The sample provides an accurate snapshot of the influence of predictor variables on scientific literacy while minimizing the impact on normal instructional activities during school hours.
3. The sample reflects the larger population for the school and other students with similar demographics.
4. School records are accurate and represent students' previous performance and demographic data.

Organization of Dissertation Chapters

The next chapter, Chapter 2, includes a review of the literature that led to the development of this study beginning with a historical overview of science education legislation and students' performance on science assessments. Science literacy is explored as a construct by understanding its impact, definition, and incorporation into K-12 curriculum. A variety of measurements for scientific literacy will be examined and critiqued. The literature review addresses the theoretical framework of this study, which was used to identify student attitudes toward science and science literacy as well as the instruments that purport to measure students' attitudes toward science. The predictor variables included in the model of this study is also defined and reviewed with justification for use.

The methodology is reported in Chapter 3 with a focus on design, instrumentation, and treatment of the data. Chapter 4 presents findings of the study and Chapter 5 includes discussion, conclusions, and implications of the findings.

Chapter 2

REVIEW OF THE LITERATURE

Introduction

This chapter begins with a brief description of the evolving role of the federal government in science education and the national and international assessments that precipitated those changes. The justification for science literacy as the goal for science education and potential impact on civil society, the economy, and academic preparation is explored. The review shifts to a historical description of science literacy, variation in definitions, inclusion in existing learning standards, and the factors that influence science literacy. The measurement of science literacy and a review of available instruments was conducted. The role of affective variables including attitude and motivation was conducted, followed by an examination of instruments that measure attitude toward science. The chapter concludes with an outline of research justifying the inclusion of variables in the science literacy model constructed.

Method of Review for Literature

The literature review was conducted using an electronic search of the database OneSearch that included e-books, EBSCO, ProQuest, JSTOR, ERIC, and Taylor & Francis Online. Results were limited to publications between 1958, which is the year of the first documented appearance of science literacy (Hurd, 1958), to the present. The search was limited to scholarly and peer-reviewed articles, books, and dissertations or theses. Search terms included: *science literacy, attitudes toward science, motivation toward science, science literacy assessment, and Test of Science Literacy*. Citations for each paper found were subsequently examined leading to additional search terms.

Historical Overview of Science Education Legislation

The U.S. Federal Government has historically played a secondary role in education because states have used the 10^{th} amendment to assert their right to govern education. The primary role of the Federal Government in public education, from conception to the 20th century, was to provide support to federal territories, access to federal lands, and initial financial support for states to develop their own educational systems. Resources were provided to higher education through the 1862 Morrill Act and the 1890 Second Morrill Act that established support for land grant universities to advance academic achievement. The Morrill Acts provided 30,0000 acres of federal land to each Representative and Senator for the creation of a public college (Second Morrill Act of 1890, §841) while ensuring equal access to land grant colleges to African Americans. Formalized federal involvement in education began with the creation of the Department of Education in 1867, which collected information and supported states' efforts to develop and foster effective educational institutions (U.S. Department of Education, 2018). Data collection began with the inception of the Department of Education to ensure allocated resources were being used effectively.

The conclusion of World War II drastically expanded the role of the Federal Government in education. During the war, the number of U.S. military personnel increased from approximately 300,000 in 1939 to over 12,000,000 in 1945. Congress passed the Servicemen's Readjustment Act (1944) to combat social unrest and support returning servicemen with financial and educational benefits. The G.I. Bill provided federal funds for returning service members to attend colleges or universities. The program resulted in nearly eight million returning service members attending colleges and universities. In 1947, nearly 50% of all students attending consisted of returning soldiers (Bound & Turner, 2002). The $4,000,000,000 increase in federal funding for education necessitated a larger federal role in oversight to prevent fraud. The institution of the Servicemen's Readjustment Act (1944) limited the federal role in education to post-secondary institutions while states governed primary and secondary education.

On October 4, 1957, the Soviet Union successfully launched the satellite Sputnik into orbit around the Earth. The satellite transmitted radio pulses as it circumnavigated the Earth, which led to press coverage that produced public panic and outrage due to a perceived deterioration of American scientific prowess (McQuaid, 2007). The fear that the Soviets were outpacing the United States in scientific and technological development was rampant with Soviets reportedly producing "two to three times as many scientists per year as the United States" (Kaiser, 2006, p. 1227). Although these numbers were later determined to be inaccurate, misrepresented, and a byproduct of Soviet propaganda, Congress acted by passing the National Defense Education Act (NDEA).

The NDEA declared the current state of education insufficient for national security needs allowing the federal government to take a broader role by linking education with national defense. The NDEA provided funds to secondary institutions, funded "key defense areas of science, mathematics, engineering, area studies, and foreign languages" (Kaiser, 2006, p. 1237), and established a student loan program to ensure "no student of ability will be denied an opportunity for higher education because of financial need" (National Defense Education Act of 1958, §85). The support to post-secondary institutions was consistent with previous federal programs, but the NDEA also included provisions for elementary and secondary schools. The NDEA dramatically increased education funding for primary and secondary schools by doubling the largest federal expenditure to date (Kaiser, 2006). The use of funding as an incentive for the adoption of policy at state and local levels remains the primary mechanism by which the federal government exerts influence over K-12 schools; elementary and secondary schools that receive federal funds are subject to federal rules and oversight.

The adoption of the NDEA in 1958 led to substantial changes in science education during the 1960's and 1970's. The NDEA charged the National Science Foundation (NSF) to establish a means of making science information widely available. The NDEA included funds for the creation of NSF curriculum material shifting the goal of science education from the acquisition of content knowledge to a scientist-in-training model (Helgeson et al., 1977). Curriculum materials funded by the NDEA were widely adopted in elementary and secondary schools with an increase in the number of students completing science courses in high school. From 1960 to 1972, the number of students completing biology, chemistry, and physics rose by approximately 1,000,000 students, 250,000 students, and 100,000 students respectively (Helgeson et al., 1977). National security concerns and emphasis on the scientist-in-training model for science education led to federal resources being spent on identifying and supporting high achieving students while doing little to address growing demographic achievement gaps (Helgeson et al., 1977).

ESEA (1965) was approved as part of President Johnson's Great Society Program. Title I of the ESEA provided federal funds to local educational agencies to support low-income students. Funding was based on average daily attendance requiring schools to track and increase attendance of low-income students for federal funding. The ESEA provided additional funding to purchase instructional materials, textbooks, and library materials. Federal money was allocated for research and distribution of education materials under Title IV of the ESEA to facilitate best educational practices. The federal government was now an arbiter in the debate over the purpose of science education and means to achieve that purpose.

The ESEA is critical legislation that is continuously updated and reauthorized. The legacy of the ESEA is the shift of federal involvement in education from national security concerns to equity. The ESEA recognized racial, gender, and economic achievement gaps as areas of concern requiring federal support to address. The ESEA sought to identify best educational practices and hold schools accountable for using those practices. The emphasis on the success of individual students escalated the importance of research and implementation of best practices. The ESEA continues to be reauthorized with political controversy over the role of the federal government in supplying resources, mandates, expectations, and definitions of best practices.

In 1981, Secretary of Education T.H. Bell created a national commission composed of 12 educational administrators, a business executive, two scientists, a retired politician, and a high school teacher to review the state of education. The commission reviewed progress toward goals established by the NDEA and ESEA as measured by the National Assessment for Educational Progress (NAEP) and College Board assessments. The findings of the commission, reported in *A Nation at Risk,* indicated that Americans' performance on standardized exams had fallen below the levels in 1956, thereby erasing all progress made in the aftermath of Sputnik (National Commission on Excellence in Education, 1983). The report cited concerns regarding U.S. students' inability to perform on international assessments, high rates of illiteracy, low academic achievement, low SAT scores, increased enrollment in college remedial courses, and consistently low science achievement scores. (National Commission on Excellence in Education, 1983).

A Nation at Risk evaluated science curriculum and implementation of science curriculum across the country. The United States was the only industrialized nation that did not require completion of biology, chemistry, physics, and advanced math for a high school diploma

(National Commission on Excellence in Education, 1983). There was a wide variation in high school graduation requirements between various states. The commission criticized science curriculum around the country as “homogenized, diluted, and diffused to the point that they [curriculum writers] no longer have a central purpose” (National Commission on Excellence in Education, 1983, p. 21). Paul Hurd summarized the findings of the commission by stating that “ae are raising a new generation of Americans that is scientifically and technologically illiterate” (National Commission on Excellence in Education, 1983, p. 12).

The commission recommended strengthening the minimum graduation requirement to include 3 years of math and 3 years of science, focusing on “methods of scientific inquiry and reasoning, [and] the application of scientific knowledge to everyday life” (National Commission on Excellence in Education, 1983, p. 33). The commission determined that an increase in instructional minutes and incorporation of new curricula focusing on methodology, literacy, and application of knowledge would lead to better educational outcomes of all students. The commission provided curricular focus by establishing science literacy as the federally supported goal of science education.

A Nation at Risk shifted discussion surrounding science education by emphasizing national deficits and spurring reform. Doubts were raised regarding the accuracy of the findings in *A Nation at Risk* and perceptions of an educational system in decline. The passage of the ESEA emphasized equity and held schools responsible for the success of all students, which resulted in a more diverse population taking SAT and NAEP assessments. At its inception, participation in the NAEP and College Board was voluntary and limited. College Board tests were initially standardized at elite preparatory schools with primarily white and economically advantaged students. Using performance assessments extrapolated using students from a small number of elite schools and applying them to the general student population nationwide was a flawed approach. The standard scores used for comparison did not represent all students. As expected, average student scores declined with the inclusion of more economically, geographically, and ethnically diverse student populations (Stedman, 1994).

The NDEA defined the purpose of science education as the development of future scientists while the ESEA promoted scientific literacy for all. These conflicting philosophical approaches of the ESEA and the NDEA, combined with local control of the curriculum, created a scattered approach to science education resulting in variance in student performance. The differing goals of the NDEA and ESEA led to significant curricular changes and approaches to science education. The NAEP Science assessment is an indicator of lagging performance that measures institutional changes implemented five to ten years earlier. The lack of progress for the general population in the 1970s resulted in numerous reform attempts in the early 1980s. During the same period, participation and performance on College Board achievement tests for biology, chemistry, and physics significantly increased, which was consistent with the goals of the NDEA.

Subsequent reauthorizations of the ESEA have incorporated many reforms that specifically target science education. In 1994, the ESEA was reauthorized under the name Improving America's Schools Act (IASA) of 1994. Elementary science resources were explicitly targeted for improvement with the establishment of the Eisenhower National Clearinghouse for Mathematics and Science Education (Improving Americas Schools Act of 1994, 2018). This clearinghouse was designed to compile, maintain, disseminate, and evaluate science and mathematics curriculum resources for both elementary and middle school students. Results from the NAEP indicated elementary students spent as little as two hours per week on science instruction even though time spent on the subject was directly correlated with performance on the NAEP (Blank, 2013). As a result, the IASA established a National Teacher Training project for early childhood in science. The IASA addressed excellence in science education but also codified expectations of equity by ensuring equal access to advances mathematics and science courses regardless of gender, ethnicity, economic need, or English proficiency (Improving Americas Schools Act of 1994, 2018). Legislation was designed to close the achievement gaps in physical science between males and females identified by the NAEP as well as between white students and racial minorities.

In 1999, a national commission to investigate the current state of American schools was established with John Glenn serving as chairman. An examination of NAEP Science scores showed continued stagnation in student performance (Campbell et al., 2000). The commission reviewed NAEP and TIMMS data before concluding the United States continues to perform below the expected standard in math and science education: “Among the 20 nations assessed in advanced mathematics and physics, none scored significantly lower than the United States and only one scored lower in physics” (National Commission on Mathematics and Science Teaching for the 21st Century, 2000, p. 11). According to NAEP Science results, only 3% of American 12th graders can be classified as having “advanced” competency in science with only 20% having achieved proficiency. Proficiency, according to the NAEP Science assessment, is consistent with the aims of science literacy requiring explanations, arguments, and an understanding of empirical evidence. The findings of the NAEP science are consistent with levels of science literacy among undergraduate students and the general public (Impey, Buxner, Antonellis, Johnson, & King, 2011).

The recommendations of the *Glenn Report* focused on improving student performance by improving the quality of instruction. The commission found that nearly half of science and mathematics teaching positions that are filled in a year are done so by either new or non-certified teachers. The *Glenn Report* recommended supporting teachers with professional development, additional resources, and changing teaching methods to support active learning. This report resulted in numerous pieces of legislation to remedy the international competitiveness of the United States, which included Goals 2000: Educate America Act; Educational Research Development, Dissemination, and Improvement Act; Carl D. Perkins Vocational and Applied Technology Education Amendments; Improving America's Schools Act; NCLB of 2001; Reauthorization Of the National Center for Education Statistics and the Creating of the Institute of Education Sciences of 2002; and ESSA of 2015.

In 2001, the United States Congress reauthorized the ESEA with NCLB, which focused on accountability. Following the recommendations of the *Glenn Report*, NCLB enacted national standards for teachers requiring that they be classified as "highly qualified" to continue to receive federal funds. A highly qualified teacher must have a bachelor's degree, full state certification, and proof of subject matter expertise in the subject they have been assigned to teach (NCLB, 2002). The State of Texas required that teaching certificate applicants obtain a bachelor's degree, complete a state approved educator preparation program, pass a Pedagogy and Professional Responsibilities exam, and pass a content knowledge exam (Texas Education Agency, 2020). NCLB included provisions that provided grants to universities, high schools, and elementary schools to foster partnerships with universities and colleges for teachers' continuous professional growth to improve the overall quality of instruction. NCLB also provided support for research and innovation in science education by fostering grant-based partnerships with universities. These partnerships proved beneficial and conducive to developing curriculum tools based on science inquiry, which resulted in significant instructional gains (Marx & Harris, 2006).

NCLB targeted math and reading assessments leading to a decline in the number of instructional minutes available for science. In 1st–4th grades, the instructional time spent on science declined from an average of 3 hours per week before the passage of NCLB to 2.3 hours per week after NCLB (Blank, 2013) with a net loss of nearly 6,000 instructional minutes during this four-year time frame. However, results from the 2005 NAEP Science showed significant gains even with decline in instructional minutes supporting a link between reading and math to science success (Grigg, Lauko, & Brockway, 2006). Further examination of the NAEP Science results demonstrates a correlation between the number of instructional hours and NAEP Science scores even when controlled for socioeconomic status. There is indication that these initial gains through improved literacy and mathematics are mitigated by the time students reach secondary school (OECD, 2016). In 2009, students beginning their education under NCLB showed significant declines in science proficiency according to the NAEP Science assessment by eighth grade (National Center for Education Statistics, 2011).

The enduring legacy of NCLB is the creation of adequate yearly progress (AYP) with mandatory assessments in content areas for promotion to the next grade. States were required to develop plans that assessed the four core areas of reading/writing, mathematics, social studies, and science. An emphasis was placed on reading and mathematics with required annual assessments for both in elementary and middle school. The heavy emphasis on reading and math resulted in the reduction of instructional minutes for science education in elementary schools, thus limiting the incorporation of high-quality but time-intensive inquiry lessons developed by NCLB partnerships with universities (Dee & Jacob, 2010).

In a 2007 report on American economic competitiveness with science, technology, and industry, the Committee on Prospering in the Global Economy of the 21st Century found that science, mathematics, and technology remain an "intense concern within the business and academic communities" (National Academy of Sciences, 2007, p. 94). The commission reviewed NAEP and International testing data, which indicated no significant progress in science and mathematics scores among U.S. students, which reduced the country's global competitiveness. Using data from the administration of the NAEP in 2000, the commission found that only 18% of American 12th graders have reached a proficient level of science understanding, which was a decline from 21% in 1996. The commission noted wide variation in scores between city, suburban, and rural schools with "some schools produc[ing] students who consistently score at the top of national and international tests while others [produce students who] consistently score at the bottom" (National Academy of Sciences, 2007, p. 98).

Rise Above the Gathering Storm focused on the economic impact of issues that may arise from a generation of students whose scores on the NAEP have consistently declined from 1996 to 2000. The number of students enrolling in and completing a science or engineering degree per capita was low compared to economic competitors. A high attrition rate in science degrees resulted in the 30% of students beginning university as a science major but only 16% earning a degree in the United States. Less than 6% of the total population in the United States earned a degree in science or engineering compared to approximately 10% for France and South Korea. Since 1985, there has been a steady decline in the number of engineering and physical/geoscience degrees awarded (National Academy of Sciences, 2007).

The Committee on Prospering in the Global Economy of the 21st Century reported a need to recruit high quality science and math majors to teach in secondary schools. NCLB began to address the need for highly qualified teachers by requiring that all teachers be highly qualified. Despite these requirements, nearly 93% of teachers teaching physical science in middle school did not major in and do not have a certification to teach physical science. The problems continue in high school where a student only has a 40% chance of having a chemistry teacher who majored in chemistry (National Academy of Sciences, 2007). The Committee on Prospering in the Global Economy of the 21st Century report advocated for the continuation and advancement of professional development for secondary science teachers. Professional development was positively correlated with continued employment and student achievement (National Academy of Sciences, 2007).

Science competency varies widely depending on the school. Variation stems from differences in curriculum, course offerings, and school funding. The Committee on Prospering in the Global Economy of the 21st Century suggested the formation of a voluntary national curriculum using 15,000 school districts independently attempting to determine best practices. Similar to other national reports, the commission found that science courses in K-12 schools "lack focus, cover too many topics, repeat material, and are implemented inconsistently" (National Academy of Sciences, 2007, p. 97). The report recommendations include the adoption of materials and standards outlined in the NSES and Project 2061. The Committee on Prospering in the Global Economy of the 21st Century concluded with a warning that "Without basic scientific literacy, adults cannot participate effectively in a world increasingly shaped by science and technology" (National Academy of Sciences, 2007, p. 112). The emphasis on science literacy and development of skills while reducing the number of content standards is consistent with the recommendations of previous commissions.

The role of the federal government has significantly shifted science education. Funding requirements have helped develop curricula, professional development partnerships, and number of science courses available in high schools. The increasing role in funding and associated mandates has promoted homogeneity in standards as evident by the Common Core State Standards (CCSS) and the Next Generation Science Standards (NGSS). Repositories of educational research and best practices are available to teachers nationwide. The purpose of general science education has evolved from training scientists to science literacy as a means of establishing civil and economic equity. Despite improved focus, funding, and attention to science achievement, student performance remains stagnant.

Historical Trends in Science Performance

The Office of Education was established in 1867 to support education in federal territories but failed to systematically measure student performance in American schools. Significant increases in funding and the expectation of improvement resulted in the creation of state, national, and international assessments to compare student academic performance and measure school effectiveness. The National Assessment of Educational Progress (NAEP), developed in the United States, was the first systematic attempt to assess and compare the performance of states and communities with national trends. International relations and the development of a global economy led to a comparison of student performance with their international peers. In 1995, U.S. students began participation in Trends in Mathematics and Science Study (TIMSS) followed by the Programme for International Student Assessment

(PISA) in 2005. The results of these assessments led to federal legislation and resources that promote science literacy as the goal for science education.

NAEP

In 1969, the Carnegie Corporation began the development of what would become the National Assessment of Educational Progress (NAEP) and the first standardized assessment to measure the student achievement of students in American schools. The NAEP was designed to support implementation of standards-based curriculum and ensure appropriate funding under the NDEA and ESEA. The first administration of the NAEP was designed to assess both content acquisition and skill development in science, writing, and citizenship. NAEP Science content knowledge requirements include life, physical, and earth science while the skills section assesses problem-solving, understanding of the nature of science, and the understanding of scientific inquiry.

Scoring was standardized across grade levels with students able to demonstrate five levels of proficiency. The first level, based on a score of 150, indicated a “rudimentary” proficiency level which included basic content knowledge. The second level, based on a score of 200, indicated a “basic” level of proficiency indicating an understanding of basic scientific principles. The third level, based on a score of 250, indicated an intermediate level of proficiency, which meant that students could apply scientific information to new situations. The fourth level, with a score of 300, was named “adept,” which meant that the student could evaluate scientific experiments and knowledge. The advanced category, based on a score of 350 or higher, was attributed to students who could make inferences and integrate specialized information across domains (Campbell et al., 1997).

The initial administration of the NAEP in 1969 resulted in a mean score for of 225 for elementary, 255 for middle school, and 305 for high school students. The average of high school students could be described as adept in science by evaluating experimental design and applying detailed knowledge to the interpretation of text and graphs (Campbell et al., 1997). The initial NAEP Science assessment was considered a pilot program used to standardize methodology. Participation in the NAEP was voluntary and a disproportionate number of students from economically advantaged, elite preparatory schools from the northeastern United States participated. Despite sampling and methodology limitations, the national high school student score of 305 became the standard to which all other assessments were compared.

The second administration of the NAEP in the 1972-73 school year resulted in substantial decline in scores for all age groups with elementary school students scoring an average of 220, middle school students achieving 250, and high school achieving 296. Direct comparison between the 1969 and 1973 assessments was not possible because of lack of common questions or standards, but the general perception was that science performance had declined. The same scoring classifications: rudimentary, basic, intermediate, adept, and advanced were used during the first administration in 1969 although the ranges were not equal due to a different framework. The 1973 NAEP Science should be considered a second pilot, useful for standardization of methodology, but not a representation of the participants' abilities (Campbell et al., 1997, p. 200).

The results of the 1977 administration of the NAEP science demonstrated results that were similar to those of the 1973 administration. Elementary and middle school scores were stable with a consistent score of 220 for elementary students and a slight decline of 3 points from 250 to 247 for middle school students. High school scores demonstrated a significant decline, down an additional six points to 290 (Campbell et al., 1997, p. 48). There was concern that despite increased federal funding through the ESEA, the United States was in danger of losing its status as an economic and technological power. A lack of common framework between 1973 and 1977 did not allow for direct comparison between the assessments; however, the decline in normalized proficiency scores was a cause for concern. The 1977 NAEP Science administration introduced a common framework for future NAEP Science assessments that allowed for longitudinal comparison for future administrations.

Student scores on the 1982 administration of the NAEP Science indicated a continued pattern of declines in performance. The mean score for high school students declined another 7 points to 283, 22 points below the initial measurement of 305. The common framework between the 1977 and 1982 NAEP Science administrations allowed for direct comparison and provided evidence that the declines in scores were not a byproduct of the assessment. Mean student performance in the 1969 administration was classified as adept*,* indicating that the average student had "detailed knowledge" and could "evaluate the appropriateness of scientific procedures" (Campbell et al., 1997, p. 24). Mean student performance in the 1982 administration was classified as *intermediate* indicating that the average student could "understand and appl[y] general information" (Campbell et al., 1997, p. 24). Disaggregation of the 1982 NAEP Science performance demonstrated that the number of students achieving each proficiency level had declined. The results of the 1982 NAEP supported the findings of the National Commission on Excellence in Education in *A Nation at Risk*.

The 1986 NAEP Science assessment was developed under a new framework specifically designed to measure national progress in math and science (Beaton & Chromy, 2010). The objective and assessment framework remained consistent from 1986 to 1996, which allowed for direct longitudinal comparison (Campbell et al., 1997). The assessment framework measured science proficiency through content, context, and cognition. This represented a small shift toward science literacy away from pure content acquisition as the goal of science education. Direct comparison cannot be conducted on items and objectives between the 1986 NAEP Science and prior assessments; however, overall proficiency levels can be compared (Mullis & Jenkins, 1988).

The 1986 NAEP science assessment resulted in higher levels of proficiency in overall performance with significant gaps in demographic comparison groups. The mean performance of high school students increased, the first increase in the history of the NAEP Science assessment, by 6 points to 289. Advanced performance continued to decline with 7.5% of students scoring a 350 or higher, a threshold for college-level readiness in science. Despite the increase in overall performance, the 1986 NAEP revealed a growing achievement gap based on high school students' race and gender. African American and Hispanic students' performance was categorized as "four years behind" their White peers while young women scored nearly 20% lower in their ability to analyze scientific procedures and data when compared to their male peers (Mullis & Jenkins, 1988, p. 10). This achievement gap was linked to the differences in attitude toward science and science literacy skills.

The 1986 NAEP Science administration surveyed the time and type of instructional activities students completed. The relative stagnation in scores nationwide was linked to low participation in science courses with only 58% of high school juniors enrolling in a science course. These courses offered less than three hours of instruction per week in science. Nearly 20% of students reported spending no time outside of class on science-related activities (Mullis & Jenkins, 1988). Increased time spent on homework was positively correlated with increased proficiency on the NAEP in high school with students spending 4 or more hours on homework scoring substantially better (average of 317) than those without (average of 285) time spent on homework (Campbell et al., 1997). The number of science courses a student completed was positively correlated with NAEP Science performance (Mullis & Jenkins, 1988).

A survey of classroom activities demonstrated that best practices in improving science literacy were limited in classroom implementation. An emphasis on content acquisition was prevalent as nearly 70% of students reported reading their textbook at least once per week in class (Mullis & Jenkins, 1988). Students did not engage with the material as most students reported never completing a lab report, reading a scientific article, or an oral or written report throughout the year (Mullis & Jenkins, 1988). The commission writing the *Nation's Report Card* in 1986 concluded that the current "content and structure of our school science curricula are generally incongruent with the ideals of the scientific enterprise" (Mullis & Jenkins, 1988, p. 20). They further argued that instead of focusing exclusively on content acquisition, science pedagogy should shift its focus toward evaluating evidence, evaluating logic, and using the tools of science to understand the world around them.

The 1990 administration of the NAEP Science continued the stagnant performance for high school students. Mean high school student performance score increased one point to 290 from 1986 (Campbell et al., 1997). There was an increase in college level readiness with advanced proficiency rising from 7.5% in 1986 to 9% in 1990. Ethnic performance achievement gaps remained with White and Asian students significantly outperforming African American and Hispanic students. Gender gaps also persisted with males scoring significantly higher than females, especially in levels of advanced proficiency. The performance gap between genders was significant with males outperforming females in life science, physical science, and earth and space science. Female students significantly outperformed males in the key subskill nature of science.

Nature of science skills is a key component of science literacy measured by the NAEP Science assessment. The NAEP Science assessment surveyed middle school teachers and found nature of science as the lowest priority in instructional objectives. Nature of science skills focus on the underlying principles of how knowledge is determined and is key to developing scientifically literate students (Jones, 1992). High school teachers had a low priority in teaching these skills with only 35% of high school teachers reporting a special priority in teaching the nature of science (Jones, 1992). Consistent with the 1986 results, the 1990 NAEP showed that the predominate instructional activity was content acquisition from the textbook with 46% of students saying they read their textbook several times per week.

Engaging with the material is key to science literacy development. There was an increase, from the 1986 to 1990 NAEP Science administrations, in the number of students reporting they read scientific articles once or more per week. Students with access to reading materials at home and a reduced time watching television had higher average proficiency scores. Discussion is critical to developing an understanding of the nature of science with 25% of students reporting they discussed science news events several times per week. This represented a shift in the expectation of involvement for students in current affairs and had the potential to make science more relevant and engaging to students. The findings of *A Nation at Risk* appeared to be implemented with greater fidelity; “if students’ science achievement [was] to improve, then school science, particularly meaningful instructional activities in science classes, need to receive additional emphasis” (Jones, 1992, p. 80).

The 1996 administration of the NAEP Science showed significant improvement in many areas. Mean high school performance increased from 290 to 296 (Campbell et al., 1997). Students demonstrated statistically significant gains in achieving higher order science skills and reaching advanced levels of proficiency. Achievement gaps were reduced with mean performance between White and African Americans students reaching a low of 47 points and the gender gap narrowing to eight points. The reduction in achievement gaps continued the trend consistent with the goals of the ESEA (Campbell et al., 1997). Overall, performance on the 1996 NAEP Science resulted in optimism but there remained some areas of concern in the survey of teachers and students. Affective variables including attitudes toward science and the value of science declined; only 24% of students believed science should be required in schools (Campbell et al., 1997).

In 2000, NAEP Science results showed an increase in the number of high achieving students and a decline in the number of low achievers. Overall, performance was consistent with the previous administration with a 1-point decrease to 295 for high school students (Campbell et al., 2000). Performance in the top 3 quartiles continued a trend of improvement from 1977 to 1999. Students in the top quartile of the 2000 NAEP Science administration outperformed their peers in all prior administrations. Overall, performance for high school students was driven by volatility in the lowest quartile of student scores (Campbell et al., 2000) with a significant decline in the percentage of high school students reaching the basic level of proficiency (O' Sullivan, Lauko, Grigg, Qian, & Zhang, 2003). The achievement gap between White students and African American students widened to 54 points in 1999 while the gap between Hispanic and White students narrowed to 30 points.

The 2005 administration was the first after the adoption of NCLB (2002). This required participation in the NAEP assessment program for receipt of federal funds, thus expanding participant schools. NCLB required schools to demonstrate progress for all students, including subpopulations based on gender, ethnicity, language, and special needs.

Mean performance on the 2005 NAEP Science remained stable with an average of 296 for high schools while certain inequities remained. Minority student performance continued a to decline compared to their 1996 scores while White students scores, compared to 2000, improved. The gender gap widened from a two-point difference in 2000 to a four-point difference in 2005. Despite the score gap, the number of male students reaching a proficient level declined while the number of female students reaching proficiency remained the same. There is no evidence that the gap between gender or ethnic groups had narrowed since 1996 (Grigg et al., 2006). Disaggregation of the results demonstrated decline in scores in areas of life science, physical science, and earth space science for high school students. The number of students graduating with a single science course decreased from 36% in 2000 to 33% in 2005. Increased enrollment in chemistry and physics did not result in gains for those areas in the NAEP Science (Grigg et al., 2006). The declines in scores should be juxtaposed against increased number of students taking multiple science courses.

The 2009 NAEP Science was developed under a new framework that used *crosscutting questions* that required the integration of knowledge from multiple subject areas. Proficiency ratings were reduced to three levels of student performance: basic, proficient, and advanced. Students classified as basic scored an average of 142 demonstrating an incomplete mastery of grade level knowledge and skills. A student scoring proficient at 179 demonstrated grade level knowledge, application of that knowledge, and analytical skills related to content knowledge. A score of 222 indicated advanced proficiency or the ability to develop alternative models, design experiments, and critically analyze investigations. The high school assessment also decreased the emphasis on earth space science to coincide with high school course selection (National Center for Education Statistics, 2011).

The mean performance on the new scale was 150, representing a 1% increase in the number of students achieving the proficient level. The gender gap widened from 2 points in 2000 to 4 points in 2005, and to 6 points in 2009 (National Center for Education Statistics, 2011). Despite a 4% increase in the number of students completing biology, chemistry, and physics courses, the number of students achieving proficiency or higher only increased by 1%. The relationship between the number of science courses completed and corresponding increases in science literacy or achievement remain unclear.

The administration of the NAEP Science assessment in 2015 demonstrated overall stagnation in scores as an area of concern. The most recent version of the NAEP Science used the same framework as the 2009 assessment with little variation in overall performance. Scores were not significantly different from the 2009 administration with the mean score for high school students increasing by one point in Life Science and Earth and Space Science (NAEP, 2015a). The achievement gap between Black and White students continued to widen in physical science, and life science, from the 2009 administration (NAEP, 2015b). The gender performance gap narrowed slightly to 5 points overall with significant variation by content discipline (NAEP, 2015a).

During the past 50 years, students' science literacy, as measured by the NAEP Science, has remained stagnant at best. Despite criticism that the decline stems from selection bias in the 1969 NAEP Science administration, scores have consistently remained below the expectations of policy makers. At conception, participation in NAEP assessments was optional for communities to assess and compare their performance with other communities, states, and national trends. The implementation of NCLB required states receiving Title I funds to participate in assessments. Local control injects significant variation into the United States educational system making the NAEP an important tool for states and communities. The passage of NCLB in 2001 eliminated the selection bias as all schools receiving federal funding were required to participate. NCLB legislation was enacted in 2001 with mean high school performance on the NAEP Science failing to result in significant improvement

Science performance, as measured by the NAEP Science, has failed to meet the expectations associated with increased attention to science curriculum, policy, and funding. An increase in the number of science courses required to graduate, focused curricula, and financial support failed to improve performance. The gender performance gap has narrowed for life sciences but remains consistent and significant for the physical sciences. The Elementary Secondary Education Act (ESEA), the NCLB, and other programs have specifically targeted ethnic performance gaps, but those gaps remain and have widened in recent administrations. There remains significant variation, nearly one proficiency scale difference, between White and African American students. The lack of progress over fifty years necessitates a shift in science curriculum focusing on the development of science literacy. The shift from content acquisition to science literacy is hallmark of the Next Generation Science Standards (NGSS). Students exposed to the NGSS will participate in the next administration of the NAEP Science scheduled for 2019. Performance on the NAEP is consistent with the findings of international assessments that measure science achievement including the TIMSS and PISA Science.

TIMSS

The International Association of the Evaluation of Educational Achievement is a nonprofit society formed in 1967 that has used variation in different school systems as experiments to find the optimal conditions for education. The International Association for the Evaluation of Educational Achievement (IEA) developed numerous assessments to compare educational systems on an international level. This led to the development and implementation of the Trends in International Mathematics and Science Study (TIMSS). In 1995, the TIMSS was administered to fourth, eighth, and high school students. The scope of the TIMSS has varied by administration but since 2003, it has included only fourth and eighth grade students.

The TIMSS was designed to test content acquisition, assess perspectives on math and science, and measure performance objectives in science. The content assessed varied by grade-level but included earth science, life science, physical sciences, and environmental issues. Consistent with NAEP Science results, elementary students in the United States demonstrated a high level of proficiency ($M = 511$) on the TIMSS scoring well above the global mean ($M = 473$). However, high school students performed ($M = 471$) significantly below the global mean ($M = 500$). An assessment focusing exclusively on physics content was of greater concern. Students in the United States had the lowest mean score ($M = 423$), which was well below the global mean ($M = 501$). The results of the TIMSS were consistent with poor performance overall and on the physical science subscale from the NAEP Science assessment.

Exploratory factor analysis of high performing schools revealed remarkable similarities across countries in key variables for students. High performing schools had students with books in the home, study aids available, parents with a high level of education, and students who spent little time working or doing chores (Mullis, Martin, Foy, & Hooper, 2016). The number of books a student has in their home had a positive impact on both science and math scores regardless of the content of those books. This indicates that economic status, parental education status, and parental involvement in education had more impact on student performance than access to educational materials. High performing schools reported a greater percentage of students completing daily homework in science, math, and other subjects (Maltese, Tai, & Fan, 2012). These findings were not consistent across countries with 79% of students reporting completing homework everyday compared to 59% of students from low performing schools reporting completion of homework. The discrepancy between homework completion and performance in the United States was among the highest in the world, indicating that assigned homework was not effective in promoting science achievement.

Subsequent administrations of the TIMSS removed high school students from the study, limiting its usefulness in establishing trends for this study. The similarity in findings between the NAEP Science and TIMSS indicates structural failures in approaches to improving the science curriculum at the secondary level. High school students in the United States perform significantly below students from other industrialized countries. The TIMSS developed methodology and standards that were partially incorporated by the Organization for Economic Development in the development of the Programme for International Student Assessment.

Programme for International Student Assessment (PISA)

In 1997, the Organization for Economic Cooperation and Development (OECD) established the Programme for International Student Assessment (PISA). The structure, methodology, and technology developed by the NAEP and enacted in the TIMSS was adopted for international comparison. The PISA is designed to assess literacy in context, differentiating between reading literacy, mathematical literacy, and science literacy. Science literacy is defined as the "capacity of students to apply knowledge and skills in key subject areas and to analyze, reason and communicate effectively" (OECD, 2016, p. 16). This definition is further divided into four aspects: knowledge, process, context, and attitudes toward science. The PISA measures process skills in science defined as the ability to explain scientific phenomena, ability to evaluate and design scientific inquiry studies, and the ability to interpret scientific data captured in the inquiry (OECD, 2016, p. 64).

In 2006, the first PISA Science was administered to fifteen-year-old students in fifty-seven countries. Of these 57 countries, 30 were members of the OECD while 27 were affiliated cooperating countries. The sample was limited to students between 15 years and 3 months to 16 year and 2 months completing a minimum of six years of formal school. Students in the United States had an average score of 489 (*SE* = 4.2), which was significantly below the global score of 500. This resulted in the United States being ranking 29th out of the 57 countries participating in the initial administration (OECD, 2007, p. 73). Scores in categories of e*xplaining phenomena scientifically* and *using scientific evidence* were significantly below the global mean with scores of 486 (*SE* 4.3) and 489 (*SE* = 5.0), respectively (OECD, 2007, p. 66). Student performance was consistent with the findings of the NAEP Science and TIMSS for secondary students.

In 2015, the PISA Science was administered with a standardized global mean of 500 (*SD* = 100). Students in the United States had an average score of 496 with a 95% confidence interval of 490-502 (OECD, 2007, p. 19), which was not statistically different from the global average 500. Disaggregation of geographic regions demonstrated significant differences with Massachusetts (*M* = 529) being among the states with the highest scores while North Carolina (*M* = 502) student scores were near the global mean. Students in the United States scored below the global mean in their ability to interpret data and evidence scientifically and explaining scientific phenomena, both of which are key aspects of science literacy. U.S. students did demonstrate higher average procedural and epistemic knowledge but lower in content acquisition in all disciplines including physical systems, living systems, and earth and space systems.

The gender gap for low achieving students, defined as scoring below Level 2, was not significantly different from the 2015 PISA Science for students in the United States. The number of low performing students decreased between 2006 and 2015 indicating a narrowing of the achievement gap for underperforming schools. There remains a significant gender gap in performance for high achieving students with nearly 10% of males reaching Level 5 or higher but only 7% of female students doing so. This gender disparity was evident in areas of epistemic beliefs including the mutability of information in textbooks, scientists' opinions, experimental findings, and scientific ideas.

The NAEP and PISA differed in the impact of the lower quartile of performance. The variance between schools in the United States was also below the OECD average of 33%, which is counter to the widely held belief that chronically underperforming schools are depressing the overall performance of American students (OECD, 2007, p. 171). Low performing schools have been linked to poverty, but the United States was among the leaders in per pupil cumulative spending with a total of $90,000 per pupil spending on students between 6 and 15 years old. This was the second highest total for all OECD, exceeded only by Switzerland. Despite this level of funding, performance was significantly below average (OECD, 2007, p. 74).

Science self-efficacy was measured with students in the United States reporting a strong belief in their ability to complete a task with little effort. The perceived abilities are not reflective of performance with U.S. students having "lower mean performance in science, but more students reporting self-efficacy in science" (OECD, 2016, p. 136). U.S. students demonstrated significantly higher motivation to learn science and future oriented motivation toward learning science (OECD, 2016, p. 147). This notable gap between self-efficacy and performance leads to unrealistic career expectations in which nearly 40% of students expect a science-related career (OECD, 2016, p. 152). The findings of the PISA are consistent with the findings of the NAEP with the U.S. falling behind its economic competitors in overall performance. A potential bright spot is the performance of high achievers; the scores of highest performing students in the U.S. were above average at nine percent and reached PISA level six while the world average was eight percent.

U.S. student scores on the PISA, collected over a decade shows that at best, the U.S. students rank average in science literacy (Fleischman, Hopstock, Pelczar, Shelley, & NCES, 2010; OECD, 2019). Despite federal and state legislation aimed to improve students' science literacy, U.S. students have routinely failed to achieve science literacy (Martin, Mullis, Foy, & Stanco, 2012). Elementary students outperform their international peers; however, those advantages are lost at the high school level, resulting in significantly lower scores by high school. Variation in state law and local control has resulted in disparity in student achievement with Massachusetts students achieving the highest scores in the world while students in Alabama score significantly below the world average.

Trends in Science Literacy

Science literacy is fundamental to a democratic society. The availability of information about science has never been greater with the wide proliferation of texts, educational television programs, entire channels devoted to science, podcasts, as well as text-based and multimedia internet sites. Miller describes a minimal level of civil scientific literacy as the "level of reading and comprehension skills needed to read the science section of the Tuesday New York Times or to watch an episode of Nova on public television" (Miller, 2012, p. 29). Although only 14% of

American adults use social media as their primary source of information, prevalence of fake news across social media platforms could influence the outcome of elections (Allcott & Gentzkow, 2017). Miller provides examples including the stem cell debate and climate change in recent election cycles to demonstrate American adult's inability to "define a stem cell" or "describe a molecule" (Miller, 2012, p. 28).

The continued mediocrity of American students does not match the science literacy of the general population. Miller found that despite the stagnant scores of American students, the science literacy of the general population grew from 10% in 1988 to 29% in 2009 (Miller, 2012). The discrepancy between the growth of science literacy in adults and students necessitates a reexamination of science curriculum including both scope and sequence.

In a twenty-year, longitudinal study of science literacy at the University of Arizona, Impey et al. (2011) used knowledge-based objective response questions with a "substantial degree of overlap with the instrument used by the National Science Foundation" (p. 35) and a 22-question Likert scale to assess scientific literacy and beliefs about science. The instrument did not report validity or reliability information beyond the assertion that the "instrument has been stable for over 20 years" (Impey et al., 2011, p. 35). The mean score for students was 11.2 out of 15 items with 7.2 out of 9 items common with the NSF assessment. The definition of science literacy for this instrument is based on a content knowledge and not on students' ability to interpret and apply information.

Schools have attempted to improve science literacy with a variety of intervention programs with varying levels of success (Majestic & Pellegrino, 2014; Moreno, Tharp, Vogt, Newell, & Burnett, 2016; Shaw et al., 2014; Walker & Li, 2016). Over the past decades, learning has shifted from home and through individual study tasks to being conducted entirely at school, thus limiting time spent on task. According to *A Nation at Risk,* the amount of homework assigned to high school seniors had decreased along with student achievement while grades continued to rise (National Commission on Excellence in Education, 1983). Impey found that students gained on average of one additional right answer by completing the two science courses required for the general education part of a non-science degree. It is notable that education majors scored "worse than average on almost all individual questions and in terms of science literacy overall" (Impey et al., 2011, p. 37).

Fifty years after the prioritization of science education and emphasis as a national goal, science literacy remains substantially below targeted goals. A variety of assessments have continuously demonstrated stagnation in science performance. The persistent stagnation in science achievement in spite of a substantial increase in financial resources, establishment of best practices, increase in the science course requirements, and social prioritization of science requires examination into factors that predict science literacy. Science literacy among all Americans remains a goal, but the lack of progress necessitates reexamination of previous approaches. Science literacy impacts society and the development of future generations through its influence on civil discourse, the economy, and high school students' academic preparation.

Justification for Science Literacy

Science achievement performance has stagnated despite an increase in resources and importance. Science and technology have been fundamental to the development of a world-wide economy, to people's perception of justice, and citizen engagement (National Commission on Mathematics and Science Teaching for the 21st Century, 2000). Improvement in science literacy performance is necessary for the continued function of the U.S. legal and democratic system as well as future economic growth. It is also essential, economy, to provide equity amid changing demographic trends.

Economic

The increasingly connected world has resulted in substantial competition for economic prosperity. The U.S. failure to prepare its students for the "health sciences and computer industries requiring science and mathematical skills" (National Commission on Mathematics and Science Teaching for the 21st Century, 2000, p. 13) has placed the country's security as well as economic at risk.

The complexity of work has dramatically increased over the past 50 years. Jobs classified as skilled labor have increased from 20% in 1950 to 85% in 2000. This is most evident in the manufacturing sector, which has transitioned from hiring largely unskilled manual laborers to relying on a highly skilled workforce of technicians and robotics experts (National Commission on Excellence in Education, 1983). The rapid pace of change requires workers to have a different set of skills compared to previous generations. The level of education required by the general populace to participate in society has drastically changed. Private industry is spending billions of dollars to train workers in skills to keep pace with a changing economic landscape.

The U.S. educational system has not kept pace with industry requiring the United States to import highly skilled workers to fill the shortage of skilled workers. The economic effect impacts both companies and the individual. An undereducated workforce limits the country's economic growth and prosperity increasing the likelihood that personal income will decline in the United States (National Academy of Sciences, 2007). The emphasis on transferable skills in science literacy supports the rapid change of pace demanded by current economic conditions.

The need to import highly skilled workers represents a national security concern as those individuals are unable to obtain security clearances necessary to maintain key components of the national infrastructure.

Investment in scientific research and development has historically spurred economic growth (Rosenberg, 1974). A substantial number of advancements are developed in partnership with federal research institutions: National Institutes of Health, National Research Council, Defense Advanced Research Projects Agency, or National Aeronautics and Space Administration (NASA). The NASA life science program has an annual operating budget of $50,000,000 but has been responsible for the development of commercial and consumer products ranging from gas chromatographs to infrared ear thermometers. The estimated economic impact of fifteen companies using technologies developed by NASA life science exceeds $1.5 billion (Hertzfeld, 2002). The Internet, originally developed by the Advanced Research Projects Agency using federal funds, was estimated to exceed $1.6 trillion in 2011 (Manyika & Roxburgy, 2011). By the first quarter of 2019, the four largest companies in the world by market capitalization provided service directly dependent on the Internet. In 2001, the cost to sequence DNA from a single individual was in excess of $100,000,000, but technologies developed through the support of the National Institutes of Health have reduced those costs to under $100 resulting in several companies making this technology available to the general public (Wetterstrand, 2016). New technologies have replaced industrialization as key economic indicators that improve the standard of living (National Commission on Mathematics and Science Teaching for the 21st Century, 2000).

Public support of science, especially funding, depends on the public understanding of the goals and processes of science (Laugksch, 2000). Science knowledge was slightly correlated with increased funding but did not have the impact that interest in science had as a predictor. The emphasis of science literacy engaging the public instead of focusing explicitly on content knowledge has the potential to positively impact funding of science programs and to spur economic growth. A lack of student participation in science can drain intellectual ability reducing the research and development capability of the country. Using the required background knowledge, the public should be able to make informed decisions about scientific issues that that impact their everyday life (Sinatra, Kienhues, & Hofer, 2014).

Civil Society

The need for an educated population extends beyond economic hegemony. A constitutional republic depends on the ability of its citizens to make informed choices about science in relation to the judicial system, elections, and public health. The judicial system requires that members of a jury weigh forensic evidence including chemical analysis, DNA analysis, and ballistic models to determine the outcome of the defendant. The public is asked to vote for politicians that support funding and establish policy for projects such as the Human Genome Project, Super Conducting Super Collider, and NASA. The democratization of information without context has the potential to negatively impact the public health of society. The ease of information access necessitates a focus on science literacy skills to provide context to information.

General citizens cannot be expected to become experts in all fields, but by focusing on science literacy, they can weigh testimonies against general principles (Thompson, 2014). For instance, the introduction of modern forensic techniques into the legal system has shifted expectations of the average citizen. It is not feasible that a jury member be an expert for a trial that may involve DNA matching, chemical trace analysis, and ballistics. Members of the jury are tasked with assessment of experts' methods used to determine guilt or innocence. Science literacy specifically addresses ability to evaluate and use information as well as the ability to solve problems critical to ensuring a fair trial.

The continued funding of large scientific projects depends on understanding and support from citizens. The Superconducting Super Collider (SSC) was scheduled to begin operation in the late 90's as the preeminent particle accelerator capable of operating at 40 GeV. The public lost faith in the project resulting in cancelation with 2 billion dollars of lost funding. Scientists had difficulty communicating the importance and potential outcomes of the SSC (Chapline, 1991). A European consortium built the Large Hadron Collider, operating at one-third the power, in 2009 resulting in the frontier of physics shifting to Europe. Cancellation of the project relegated the United States to a minor role in this frontier of experimental physics. The cancellation of the SSC reminded scientists of the importance in communicating with the public and resulted in efforts to support science literacy.

Scientific information has become readily accessible, which has led to a democratization of expertise among the general public. The rise in the number of measles cases in the last decade has been linked to the proliferation of falsified research papers via social media (Dredze, Broniatowski, Smith, & Hilyard, 2016). This led to a public health crisis affecting populations in major urban centers from a disease believed to have been eradicated. The debate over the existence of anthropogenic global warming continues despite overwhelming evidence. Vaccination and global warming represent significant health risks for not only individuals who may not believe scientific facts, but they also threaten the public at large. The structure of a democracy and accessibility of information necessitate a shift from content acquisition to critical evaluation of available information.

Changing Demographics

The population of the United States is estimated to increase by 20% between 2005 and 2030 (National Academy of Sciences, 2007). African American and Hispanic populations are growing faster than other ethnic groups. These groups have historically struggled to achieve the minimum standard of literacy required to participate fully in society, as measured by the National Adult Literacy Survey (National Academy of Sciences, 2007). This has the potential to profoundly impact the economic, social, and educational outcomes of the United States as a whole. The increase in population due to immigration and higher birth rates indicate that without intervention, the average literacy scores in the United States will continue to decline. The historic inequity of ethnic performance gaps represents a systemic issue.

Despite the increase in courses and additional allocation of resources, no significant progress in educational attainment has been made. The percentage of high school students completing science courses in Biology, Chemistry and Physics has increased by 5%, 21% and 15% respectively from 1990 to 2009. Students earning credit for all three courses increased from 19% to 30% during the same period. There was a reduction in student performance achievement gap between ethnic groups during this time. Nearly three times the number of Asian students complete all three science courses when compared to African Americans. However, according to the OECD, the increase in the number of science courses completed during this period did not result in improvement in scientific knowledge (OECD, 2000; OECD, 2016).

Texas is at the forefront of demographic changes. Failure to proactively demographics changes led to a decline in student performance relative to the world. The OECD ranked Texas as 24th in the world based on the percentage of its population that holds an associate or bachelor's degree, which has been declining significantly during the past 30 years (OECD, 2016). As of 2018, just 32.1% of adults from age 25-44 in Texas had a bachelor's degree or higher while the national average is 36.4%; in Massachusetts, the average is 52.4% (National Science Board, 2020). Failure to adequately prepare students has led to a significantly less educated workforce.

The increase in the number of science and math courses completed has failed to prepare students for post-secondary education. The Community College Resource Center found 51% of students require some form of remediation (U.S. Department of Education, 2017). Students entering the remedial pathway are less likely to graduate. In a 2015 report on the state of college readiness, only 38% of students reported being college-ready in science (ACT, 2015). College readiness in science has significant ethnic performance gaps with 70% of Asians becoming college ready but only 14% of African Americans doing so (ACT, 2015). Declining post-secondary performance must be juxtaposed with higher graduation rates. The high school graduation rate rose from 83.7% in 1973 to 92.0% in 2013. The rising high school graduation rates and decreased college readiness highlights a disconnect between the knowledge and skills being taught at the high school level with those expected at the collegiate level.

Science Literacy

The continued function of a democratic society and economic prosperity require a shift in educational objectives to focus on science literacy development. In 1958, Paul DeHart Hurd (1958) defined science literacy as the goal for all Americans but failed to provide an operational definition. The definition of science literacy oscillates between extreme variation in the role of social influence, content knowledge, and nature of science (NOS) skills, which has resulted in a continuum of definitions (Roberts, 2007). Definitions range from specified content acquisition to the social relationship between communities and science. Scientific literacy is excluded as a term as it “refers to literacy that is scientifically sound no matter what content domain it focuses on” (Roberts, 2007, p. 4).

This section provides a brief overview of literacy, not science literacy, as a generalized concept including components that are included in a subset of science literacy definitions. Science literacy includes component pieces of other types of literacy including generalized literacy skills, informational literacy, and quantitative literacy. Variation in the definition of science literacy is explored including hierarchical and continuum approaches. The term *science literacy* is used instead of scientifically literate as literate implies a finished product and is strongly associated with the ability to read and write (Miller, 1983).

Generalized Literacy

A universal definition of general literacy is nebulous due to the different interpretations of literacy as a skill or social practices. In the *Education for All A Global Monitoring Report*, the United Nations describes the historical origin of literacy as a state of being “familiar with literature” or broadly as “well-educated” (United Nations Education Scientific and Cultural Organization, 2006). The industrialization of the economy at the beginning of the twentieth century necessitated an educated workforce. This need led to changes in the educational system including a focus on reading, writing, and basic numeracy. The 1930 Census defined illiterate individuals as those ten years or older that could not read and write in any language (U.S. Department of Commerce, 1979). The definition was later adjusted to functionally illiterate in which a person has less than 5 years of formal schooling (United Nations Education Scientific and Cultural Organization, 2006).

In 1947, literacy was defined as the ability to read and write at a basic level. This definition remained fairly consistent for nearly seventy years (United Nations Education

Scientific and Cultural Organization, 2006). The number of people who were unable to read or write decreased from 20% in 1870 to 0.6% in 1979 (United Nations Education Scientific and Cultural Organization, 2006). The definition of literacy has evolved with reading and writing at the core of operational definitions. The National Assessment of Adult Literacy defines literacy as the ability to use “printed and written information to function in society, to achieve one's goals, and to develop one's knowledge and potential” (Kutner et al., 2007). The definition of literacy has evolved beyond reading, writing, and numeracy to reflect a set of tools for the acquisition of knowledge and the enhancement of critical thinking rather than a set of goals and ends unto themselves (National Research Council, 2014).

The shift of educational goals from generalized literacy to content knowledge acquisition has resulted in a lack of higher performance for upper grade levels. Generalized literacy skills are the major curricular focus in elementary schools resulting in competitive global performance as measured by the early grade PISA assessments (OECD, 2016). The lack of focus at the secondary level has produced a culture of apathy among teachers and students regarding adolescent literacy. The lack of curricular focus on adolescent literacy has resulted in only 5% of high school students reaching advanced levels of literacy (Vacca & Alvermann, 1998). High school students would benefit from explicit instruction in literacy as the length, type of text, vocabulary, and structure of the text change within each discipline at the secondary level (Biancarosa & Snow, 2004). These generalized literacy skills can directly support the development of science literacy skills (Aberšek & Aberšek, 2013). Higher levels of basic general literacy skills are correlated with professional and managerial positions, lower levels of poverty, and increased civic participation (National Academy of Science, 2007).

In a synthesis of the literature, Tang (2015) noted the limitations of defining literacy in societal terms when measuring it in a multicultural society. The shift from reading and writing to tools for the acquisition of knowledge require the definition of literacy to shift based on discipline. The definition of literacy was so broad that it could no longer be operationally defined. Multiple literacies, including traditional literacy, informational literacy, and quantitative literacy arose as operational definitions for specific investigations. Although the concept of multiple literacies has prevailed in recent decades, detractors criticize the lack of emphasis on reading and writing in favor of multiple literacies (United Nations Education Scientific and Cultural Organization, 2006). The lack of development of generalized literacy skills has been linked to poor performance in science literacy assessments (OECD, 2016).

Information literacy. Informational literacy is the way an individual interacts with new information and is directly related to science literacy. The Association of College and Research Libraries defines informational literacy as the ability to recognize the need for information, access that information, evaluate the information, and effectively use that information (American Library Association, 2000). A national survey of students by the Kaiser Foundation found that adolescents spend an average of nearly 8 hours per day consuming media as opposed to 38 minutes interacting with print (Rideout, Foehr, & Roberts, 2010). Informational literacy differs from generalized literacy in that it focuses on nonfiction texts.

The emphasis on reading standards in the NCLB has resulted in an increase in the amount of instructional time spend on reading with most of that time spent on fictional texts. The preference for adolescents to engage with fictional texts may stem from the training received in early grades (Milliot, 2012). In a survey of first graders, Duke found that students spent an average of 3.6 minutes per day interacting with informational texts and informational texts represented only 6-11% of available texts (Duke, 2000). The time spent interacting with nonfiction texts declined from 2000 to 2009 with students reporting an average of 12 minutes per day interacting with traditional nonfiction texts such as newspapers and magazines outside of schoolwork.

Science literacy requires students to interact with informational text and to evaluate the validity of source material, which overlap with informational literacy skills. A survey of Australian high school students who had received OECD scores similar to students in the United States found that most teenagers struggle to identify, evaluate, and make use of academic research, instead relying on search engines (Salisbury & Karamanis, 2011). Informational literacy represents two of the nine skills assessed by the Test of Science Literacy Skills assessment requiring students to evaluate scientific information and the validity of sources (Gormally et al., 2012)

Quantitative literacy. Science literacy is tied to quantitative literacy as information is commonly reported in numerical, tabular, and graphical terms. The National Center for Education and Statistics defines quantitative literacy as knowledge and skills required to perform calculations related to everyday life activities (Kirsch, Jungeblut, Jenkins, & Kolstad, 2002; Kunter et al., 2007). The ability to interpret, analyze, and communicate scientific conclusions creates significant overlap between definitions of science literacy and quantitative literacy.

Performance on science literacy assessments correlates with performance on quantitative literacy performance. Quantitative literacy performance had slight gains 1992 to 2003 with a narrowing of the gender gap similar to NAEP Science results (Kunter et al., 2007). There are significant but narrowing ethnic performance gaps with a 66-point difference between White and Hispanic adults (Kunter et al., 2007). This trend is largely mitigated when accounting for

Spanish speaking homes, indicating a substantial relationship between language and quantitative literacy. Quantitative literacy represents two of the nine skills assessed by the Test of Science Literacy Skills assessment requiring that students understand statistics and perform calculations (Gormally et al., 2012).

Goal of Science Education

Prior to the twentieth century, science education was relegated well below the classical humanities in educational priority (DeBoer, 2000). The inclusion of science in the general curriculum resulted in a discussion about the purpose of science education. States, colleges, and schools debated whether to focus science education on students' acquisition of content knowledge, becoming scientists in training, or understanding the interaction between science and society. In 1918, the NEA proposed that science education focus on the "application of knowledge to the activities of life, rather than primarily in terms of the demands of any subject" for the general citizen (Department of the Interior, 1918, p. 8). The goals of the National Defense Education Act (NDEA) shifted the focus of science education from the general development of the citizen to the development of scientists (Kaiser, 2006). The emphasis on science education for everyday life has returned with an emphasis on the general citizens' ability to use scientific reasoning to engage in and make decisions regarding vaccines, astrology, and global warming (Rudolph, 2014). Today, science education helps students understand the natural world and as well as how scientific knowledge applies to issues related to citizenship (DeBoer, 2000).

The detonation of atomic weapons at the conclusion of World War II created a juxtaposition of pride in American preeminence while exacerbating concern over the potential harm of caused by scientific advancement. The nuclear capabilities, ballistic missiles, and advanced aeronautics of the United States and Soviet Union now had the capability of altering or even destroying society. In response, curriculum shifted away from the development of scientists to the relationship between society and science (Kaiser, 2006).

The launch of Sputnik resulted in a shift of the curriculum back to the development of scientists and away from science for the general citizen. The National Science Foundation developed a curriculum focused on the development of scientists and citizens that were supportive of the effort of scientists (DeBoer, 2000). Public support for science education soared after the launch of Sputnik and continues to remain high with 79% of adults agreeing that science has had a positive impact on their lives (Funk, Rainie, & Page, 2015). The relationship between the goal of science education, society, and content knowledge continue to oscillate.

Curricular and policy changes for science education have changed the priority between scientist preparation and applying scientific knowledge for the average citizen many times since the conclusion of World War II. This repetitive pattern has been continuous from the writings of Dewey in 1909 through the development of the Next Generation Science Standards (NGSS) in 2013. A key difference between the two approaches is a focus on the production of or consumption of scientific information (Millar, 2002). Science literacy has the capability of bridging the divide between production and consumption by focusing on the development of critical thinking skills applicable to both, the average citizen making decisions or a scientist developing experimental protocols.

Origin and Evolution of Science Literacy

The definition of science literacy in research continues to fluctuate with the priorities of society. This section explains the historical origin of science literacy examined with links to legislative change. The splintering of science literacy into hierarchal definitions from cultural science literacy to true science literacy are reviewed in order of cognitive expectation. Aggregate definitions and classification systems are examined with a focus on efforts to categorize and unify the various definitions of science literacy culminating in Roberts classification of science literacy as a continuum between two competing visions. Vison I, which supports the development of scientific knowledge, and Vision II, which emphasizes the scientific method (Roberts, 2007).

When science literacy was first conceptualized in 1958, it was synonymous with basic scientific knowledge that all citizens should have regardless of their career path (DeBoer, 2000). Hurd (1958), in his first documented use of the term, advocated an understanding of science that included a small percentage of overall known content, which would allow the learner to gain depth of understanding (Hurd, 1958). However, Seitz (1958) argued that science curriculum should provide society with a tangible benefit by emphasizing accomplishments, goals, and the impact of the field of science on society (as cited in DeBoer, 2000). This view could relegate science education to a propaganda tool for the expansion of science.

Science education should help society develop a realistic view of what science can accomplish while capitalizing on students' interest and wonder to support social scientific endeavors. Early debates over defining science literacy centered on the quantity of content knowledge and identification of content knowledge for a citizen to be scientifically literate. Johnson (1962) differentiated between the citizen and the future scientist using the concept of science literacy. A scientifically literate individual would be "conversant with the ideas that are being considered in the intellectual marketplaces around the world" (Johnson, 1962, p. 259). This distinction led to the first operational definition of science literacy that included content knowledge, research methodology, and attitudes toward science. These three domains of science literacy are present, in varying degrees, in current iterations of science literacy definitions.

By the early 1960s, implementation of NDEA in schools across the country in response to national security concerns had already occurred. The new course curriculum emphasized academic rigor to develop the most gifted students. Content experts training the next generation of scientists developed a curriculum that was packed with content and inconsistent with the cognitive development of high school students. A survey of scientists by the National Science Teachers Association (NSTA) found that science literacy in high schools developing a thorough content knowledge in multiple disciplines without including the interaction of science and society (Noddings, 1992).

The curricular reforms of ESEA (1965) prioritized equity and equality through education over national defense. This shift reduced the emphasis on content knowledge in favor of developing secondary students' understanding of the relationship between science and society. Hurd (1970) concurred with the approach favoring scientifically literate population as the only way to progress in educational aims (Deboer, 2000). The structural changes of the ESEA resulted in a shift away from the scientist-in-training model to the development of a scientifically literate population that could apply the principles of science in everyday life. NAEP Science performance declined after the implementation of the ESEA although it has not been quantified to what degree the shift in priorities impacted performance declines

Definitions of Science Literacy

Within a decade of Hurd (1958) proposing science literacy as the goal for all Americans, numerous definitions for science literacy developed. Early definitions of science literacy focused on two major dimensions of science literacy: the norms of science and content knowledge from various disciplines (Miller, 1983). The variety of definitions of science literacy produced a schism in the educational research community and a failure to develop a universal, operational definition for science literacy. Variations in definitions of science literacy developed based on students' age, educational goal, teachers' pedagogical approach, and definitions of measurable outcomes. Consensus emerged as science literacy was defined as a more complex construct than the minimum understanding required to function in society (Laugksch, 2000). Definitions of science literacy were divided into sub classifications representing both equivalent and hierarchal classification systems. Variations in the definition of science literacy included civic, cultural, practical, functional, true, science and technological, nominal, conceptual and procedural, sociocentric, and multidimensional (Roberts, 2007). There remains a lack of consensus over a universally applicable operational definition for science literacy. The following section explores the various approaches to science literacy followed by attempts to aggregate definitions into a classification schema.

Civic science literacy. Civic science literacy requires one to have background knowledge to participate in and understand public debates related to the effects of science on society. This level of science literacy does not include the ability to conduct or comprehend research used to generate scientific knowledge (Hazen & Trefil, 1991). Civic science relates to the role that individual had on society by voting on science-related issues and using the processes of scientific inquiry to make wise personal decisions. A key component of civic science literacy is differentiation between personal opinions and scientific understanding to improve personal decision-making and civic engagement. The cognitive level of development associated with civic science literacy depends on the level of expected engagement in society.

Durant (1994) defined science literacy as the reasonable knowledge a citizen must have to function in modern society. The focus on functioning in society instead of engaging in the discourse represents a lower cognitive expectation for science literacy. An approach that focuses on awareness and understanding of science that allows an individual to "bring common sense to bear" and "participate more fully in the democratic process" has a lower cognitive demand when compared to engaging in civil discourse (Shen, 1975, p. 266).

Miller defined civic science literacy as a three-dimensional construct related to content knowledge, the nature of science, and the relationship between science and society (Miller, 1983). Miller later removed the relationship between science and society because of lack of evidence to support its inclusion. Civic science literacy requires an understanding that allows an individual to engage in civic discourse of scientific issues and individualized decision-making. Miller defined civic science literacy as the knowledge and ability required to understand news reports and converse about science (Miller, 1998). This definition is problematic as a universal operational definition because the capacity to read and converse about publications depends on language proficiency, knowledge of domain-specific terminology, and understanding of domain-specific concepts.

The second competency outlined by the Organization for Economic Cooperation and Development (OECD) in the development of the PISA includes civil scientific literacy as the ability to visualize data, establish patterns, interpret information, and use empirical evidence to justify or reject claims (OECD, 2007). The OECD definition is more universal as it includes the process and skills associated with science instead of content knowledge and understanding of terminology associated with reading and conversing about science. The ability to use empirical evidence to evaluate claims is a fundamental skill critical for individuals' participation in a democratic society. Civic science literacy focuses on the process of reading, interpreting, and writing with little domain-specific content knowledge.

Cultural science literacy. Cultural science literacy incorporates the historical development of science and interaction between science and society into the definition of science literacy. Shen defined cultural science literacy as an understanding of how science is a human creation used to explore the natural world (Shen, 1975). Shen's view of science literacy suggests that science can be as appreciated as art as a form of reflection of human knowledge and society. Minimal content or methodological knowledge is necessary with a focus on the relationship between science and society. A cultural approach would fundamentally transform science education creating separate tracks for science appreciation and scientist in-training models. The appreciation of science and nature has been included in several aggregate definitions of science literacy (Layton, Jenkins, & Donnelly, 1994) but a shift toward science appreciation has not taken hold.

Cultural science literacy has been defined as one that requires the lowest cognitive demand in hierarchal approaches to science literacy (Shamos, 1995). Shamos (1995) suggested that this level represented the common background knowledge that individuals have when communicating about science). This cultural approach requires a lower level of discipline-specific terminology and conceptual understanding. A key difference between Shamos and Shen's definitions is the application of science information to the human experience or matters that include science.

Bybee (1997) has a similar description for science literacy using the term *nominal science literacy* as a substitute for cultural science literacy. Nominal science literacy is when an individual has limited understanding of terms and lacks an understanding of the underlying concepts of science literacy (Bybee, 1997). This definition overlaps with other cultural approaches by focusing on limited content knowledge without an understanding of the methods to obtain that knowledge. A common factor in the approaches of nominal and cultural science literacy is an emphasis on shared background knowledge independent of understanding the methods used to obtain that knowledge.

Practical science literacy. Practical science literacy focused on the tangible benefits that understanding science can have on the health and safety of communities and individuals. Shen (1975) defined practical scientific literacy as "scientific and technical know-how to immediately put to use to help improve living standards" (p. 265). Practical science literacy focuses on solving problems that are highly specific to individuals or communities

Practical science literacy introduced a social framework for measuring the success of practical science literacy. An individual must be successful in their specific setting to achieve practical science literacy. This results in a high degree of variation from community to community and prevents meaningful cross-national assessment. For instance, health concerns in industrialized nations may focus on limiting caloric intake and exercise while developing countries may focus on preventative measures to avoid infectious diseases spread through water sources. Both issues are critically important to communities in which the individuals exist, but the nature of the problems prevent meaningful comparison at the state, national, or international levels.

The use of conceptual science literacy outlined by Bybee (1997) corresponds to the cognitive requirements of practical science literacy. Individuals reaching conceptual science literacy understand the relationship between individual science disciplines such as biology, chemistry, and physics with an incomplete understanding of how that knowledge is produced

(Bybee, 1997). Bybee (1997) and Shen's (1975) approaches to practical science literacy require that individuals understand scientific content and methodology to reach a specific level of science literacy. Shen focused on solving problems at the individual level while Bybee emphasized social or collective decision-making. In a hierarchal approach, conceptual literacy is the lowest level of science literacy that requires an understanding of scientific methodology referred to as nature of science skills (NOS).

Functional science literacy. Functional science literacy is with the mid-tier in the hierarchy of cognitive expectations for science literacy (Shamos, 1995). Functional science literacy closely aligns to generalized literacy because it focuses on reading and writing using discipline-specific concepts and vocabulary (Bybee, 1997). This type of science literacy can be measured by assessing individuals' ability to communicate effectively by using science information but without technical expertise. Tourney (2010) approached functional science literacy as "science in the service of citizens and consumers" (Allchin et al., 2014, p. 462). This definition requires limited knowledge of NOS skills in a content neutral setting to allow for generalized applicability to consumer settings.

A broad definition of functional science literacy is the knowledge individuals need for the setting in which they find themselves or the specific knowledge that someone must know for the setting they are in (Ryder, 2001). Functional science literacy can be classified as a situationally dependent understanding of science literacy. The goal of functional science literacy is similar to practical science literacy because both focus on developing students' understanding of the methods used to develop scientific knowledge (Shen, 1975). NOS, critical to functional science literacy, can be influenced by the inclusion of student inquiry, historical cases, and contemporary cases (Allchin et al., 2014).

True/multidimensional science literacy. True science literacy has the highest cognitive expectation of all the definitions of science literacy. The highest form of science literacy in a hierarchal approach includes a complete understanding of the conceptual structure of science and the historical development of science (Bybee, 1997). This requires that an individual obtain an understanding of broad scientific concepts, specific content knowledge, knowledge of creation techniques, the nature of science skills, the conceptual linkage between domains and disciplines, and deductive and inductive reasoning (Shamos, 1995). It is estimated that only 7% of the U.S. population has reached this highest level of science literacy (Roberts, 2007).

True science literacy requires proficiency or mastery of multiple component literacies leading to the more commonly used term of multidimensional science literacy. Multiple literacies have been included in definitions of science literacy since the inception of the term in 1959, when C.P. Snow suggested that the goal of science literacy should be to support a cross-curricular approach linking science education to the humanities (Miller, 1983). The origin of science is a methodology for approaching philosophical questions the humanities had yet to logically answer. Despite their common historical origins, the humanities and sciences have largely grown apart. Science curriculum has not included general literacy principles used to acquire and communicate scientific ideas (Liu, 2009). Students are expected to read multiple texts, communicate using discipline-specific vocabulary, and use expository writing to convey concepts and ideas.

The American Association for the Advancement of Science (AAAS) defines science literacy as a subject that includes an understanding of science, math, and technology (American Association for the Advancement of Science, 1990). Scientific concepts require multi-modal explanations using verbal, mathematical, graphic, and procedural knowledge (Lemke, 1998).

Proficient and advanced students must be experts in their field as well as the tools of other scientific disciplines (Tang, 2015).

Science literacy requires transcendence of discipline-specific boundaries for a deeper understanding of concepts using multiple reference frames (Laugksch, 2000). Miller proposed that multidimensional science literacy must include NOS, key terms and concepts, and the relationship between society, science, and technology. Factor analysis led him to reject the inclusion of a social component to science literacy but that may have been due to mediation of the social component by attitude toward science (Laugksch, 2000).

Science literacy and informational literacy are intertwined and include overlapping skills. Jarman and McClune (2002) sought to incorporate the interpretation of scientific knowledge conveyed in newspapers into the definition of science literacy because newspapers are the primary source of science information for adults (Jarman & McClune, 2002). Science literacy requires “knowledge, processing skills, and disposition” required to interact with science information (Murti, Aminah, & Harjana, 2018). This requires the inclusion of critical, media, and informational literacy into the definition of science literacy.

Science literacy includes the training of the general population in scientific methodology. Although critical to the continued development of society, a training program for scientists extends beyond the knowledge and skills of the average citizen. However, Shamos (1995) argued that it was “naive to think that . . . students can learn to think like scientists” just as it would be naïve to assume that the general populace could become experts in other fields (Shamos, 1995). Science literacy incorporates multiple literacies integrating them as part of an interaction with scientific content.

Sociocentric science literacy. Sociocentric definitions for science literacy significantly differ from others in that an individual cannot be scientifically literate or illiterate. Sociocentric science literacy requires consideration of the interaction of scientific knowledge in a social context. Science literacy is not a standard that can be attained by an individual but reflects the evolving relationship between science, individuals, and society. A sociocentric approach to science literacy includes the interaction of school science education, extracurricular science knowledge, and the expectations of society in a continuous, lifelong improvement model for science literacy (Liu, 2009). This model cannot conform to the rigidity or uniformity of a nationalized curriculum but should respond to local concerns in a manner like practical science literacy. The local response aspect presents difficulty in assessment and comparison. Sociocentric science literacy is interchangeably used with science for specific social purposes and includes the interaction of science knowledge, local knowledge, and knowledge in other forms to address a specific problem in that community (Roberts, 2007). Sociocentric science literacy matches with generalized literacy that knowledge and skills are developed for social purposes.

Holbrook and Rannikmae (2007) developed the most comprehensive and measurable definition for science literacy as the "ability, to creatively utilize appropriate evidence-based scientific knowledge and skills, particularly with relevance for everyday life and a career, in solving personally challenging yet meaningful scientific problems as well as making, responsible socio-scientific decisions" (Holbrook & Rannikmae, 2007, p. 286). Evidence-based scientific knowledge and skills can be operationally defined in learning standards allowing local control over the life and personal application of those standards. This sociocentric approach to science literacy can be measured through observation of argumentation (Cavagnetto, 2010) at the participant level and aggregated to reflect society at large. Deboer (2000) has suggested that the nebulous nature of the definition of science literacy provides local schools the opportunity to pursue goals suitable to their community, which conflicts with the general pattern of nationally normed science assessments.

Sociocentric science literacy has no value as an individualized construct, instead defining science literacy as a reflection of society (Roth & Lee, 2002). Individualized scores on science literacy in a sociocentric approach show societal trends. This approach conflicts with the original goal that all Americans become science literate. Sociocentric science literacy also includes analysis of various constituencies interested in science education and includes schools, zoos, museums, policymakers, and companies with political clout. This interpretation suggests that science literacy is a means of maintaining power structures within society by providing knowledge to some groups, emphasizing the importance of this knowledge, and depriving other groups of the opportunity to pursue this knowledge (Bailey, 1998).

Content Dependency

The operational definition of science literacy and its relationship to content knowledge has evolved over time. An aggregation of definitions of science literacy reveals uniformity in the idea that science literacy includes content knowledge, but there is significant variation in the identity and degree of that knowledge (Gabel, 1976). The greater the degree of content dependency in the operational definition of science literacy, the greater the emphasis on a deficit model. The lack of a concrete operational definition bears the risk that science literacy may become obsolete and a relic of an educational fad (Miller, 1998).

Content included in science literacy is broadly defined as “what the public should know about science in order to live more effectively with respect to the natural world” (DeBoer, 2000,

p. 594). This is highly subjective and depends on current events and constituencies. Scientists have historically influenced the selection of content knowledge in learning standards (Roberts, 2007). Scientists that developed the curriculum for the National Science Foundation (NSF) had a vested interest in ensuring that their content disciplines are included as key components within the learning standards. Multiple constituencies created a hodgepodge of learning standards that do not meet the cognitive and social needs of students.

According to National Science Education Standards, science literacy is "broad and includes virtually all of the objectives of science education that have been identified over the years" (Deboer, 2000, p. 590). This does not represent a realistic learning goal for students based on either the scientist-in-training approach or average citizen approach. The inclusion of content knowledge in the definition of science literacy could result in a continually shifting target that progressively increases in difficulty as scientific knowledge becomes increasingly complex. For example, fifty years ago, DNA was not included in the general biology curriculum because it represented cutting-edge research, but now it is considered essential and is included in the Texas Essential Knowledge and Skills for Science (2018) seven times.

In recent years, there has been a shift toward considering science literacy content neutral and focusing more on students' understanding of the process and application of scientific principles. Big conceptual understanding of content is not a critical element of science literacy but instead reflects the context through which science literacy can be assessed (Holbrook & Rannikmae, 2007). The shift to nature of science (NOS) definition is advantageous because its focuses on the scientific methods allow measurement of science literacy regardless of the scientific content presented in a course. The generalizability of skills that make up a NOS definition of science literacy can allow students to transfer their skills to new and unresolved scientific questions without first accumulating massive amounts of content knowledge.

The OECD includes science content knowledge, scientific processes, and scientific situations in the assessment of student science performance (OECD, 2007). The most recent framework used to develop PISA Science assessments defined science literacy as "the ability to engage with science-related issues, and with the ideas of science, as a reflective citizen" (OECD, 2017, p. 75). The framework for the PISA requires students to interpret data, explain phenomena, and evaluate experimental designs. Success on the first domain of the PISA depends on the acquisition of content knowledge. The OECD definition of scientific literacy represents a summation of all potential aspects of science education and is subject to changing course sequence and content variation within localities. The national and international expectation that content knowledge is a fraction of student expectations necessitates a reevaluation of current learning objectives and assessments. The public should be able to make informed decisions about science that relate to their everyday life while using the background knowledge necessary to impact those decisions (Sinatra et al., 2014).

Aggregate Definitions

Science literacy exists as a broad, generic term under which numerous operational definitions exist. From 1958 to 1966, hundreds of different approaches to science literacy had created confusion about the term. Pella (1966) reviewed over 100 papers to create an aggregate definition of science literacy by categorizing the components of each definition. The most common factors included in each definition included basic content knowledge, nature of science skills, ethics, and differentiating between science and technology, science and the humanities, and science and society (Pella, 1966). Science literacy definitions could be scaled on the inclusion and ranking of these six factors in the operational definition.

Gabel (1976) refined Pella's work by developing eight dimensions of science literacy and creating a matrix to classify different definitions. One side of the matrix included cognitive factors while the other had affective factors. The cognitive domains included the differing levels of Bloom's taxonomy while the affective domains included valuing, behaving, and advocating. Academic literature that defined science literacy in advocating the interaction of science, technology, and society was placed in that element of the matrix. This analysis of existing definitions was able to demonstrate an entry into nearly every element of the matrix indicating that the term scientific literacy had no clear definition and was unlikely to be consolidated into a universal definition (Gabel, 1976). This use of cognitive and affective factors is consistent with modern approaches to science literacy while demonstrating continued fracturing.

An aggregate definition of science literacy can be triangulated by examining factors used in the development of the definition. Laugksch (2000) created a conceptual framework for science literacy classifying uses in terms of interest groups, number of attributes, level of mastery, justifications for, and measurement of science literacy. Scientists, science educators, and policymakers all approached science literacy which can impact the approach. The number of attributes included in defining science literacy range from 3 to 12 with emphasis ranging from pure content to the interaction of society with science. Attempts at hierarchal expectations for science literacy create a range from minimum to function to advanced. The NAEP Science and PISA use similar hierarchal expectations with the NAEP using basic, proficient, and advanced while the PISA operates on a numerical tiered scale of 1 through 5. The justification for science literacy will affect the operational definition with the most common justifications being increasing democratic participation, national defense, and economic security. Measurements of science literacy vary from surveys of the public to measurement of social science literacy to psychometrically validated assessments measuring the science literacy of an individual. A universal definition of science literacy is intractable, so all operational definitions need to be compared using Gabel's matrix or framework of Laugksch.

According to the Public Understanding of Science (PUS) initiative, universal science literacy reflects society's key interests (Shen, 1975). Citizens are often required to select candidates, vote on resolutions, and make lifestyle choices that require fundamental knowledge of science. The wide-ranging scope of science make it improbable that citizens can become experts in all fields; instead, they rely on the process of science and science literacy. Using the Laugksch (2000) classification approach, the PUS definition of science is based on the perceived need to improve civil and governmental affairs, so it focuses on citizens' understanding of the scientific process so that they can function in society.

The American Association for the Advancement of Science (AAAS) began Project 2061 in direct response to *A Nation at Risk* with a focus on developing science, technology, and mathematics literacy among all Americans (American Association for the Advancement of Science, 1990). This group sought to increase the scientific literacy of the general population by reducing the amount of content taught in schools and focusing on a common core of ideas and skills to maximize scientific literacy (American Association for the Advancement of Science, 1990). The AAAS included both key content knowledge and NOS skills in the atlas of science literacy outlined in Project 2061 (Bybee & McInerney, 1995).

According to the AAAS, science literacy includes content knowledge, procedural knowledge, and affective variables related to science, math, and technology. The goal of Project

2061 was to mitigate declining science achievement scores among K-12 students and to promote science literacy for all Americans. Though never properly implemented, Project 2061 has contributed to the science literacy debate and has resulted in the development of new curricular materials. The AAAS is a private, non-profit organization dedicated to advancing science, so it did not have the means to effectively advocate for the adoption of its curriculum in all 50 states. However, key concepts and ideas from Project 2061 were included in the Next Generation Science Standards in an effort to establish a national science curriculum (American Association for the Advancement of Science, 1990; Next Generation Science Standards Lead States, 2013). The National Research Council, in conjunction with the National Science Foundation, National Aeronautics and Space Administration, National Institutes of Health, National Academy of Sciences, and U.S. Department of Education developed a series of science educational standards referred to as the National Science Education Standards (NSES). The purpose the NSES was to "achieve the goal of scientific literacy for all students that is described in the content standards" (National Research Council, 1996, p. 1), which included understanding the nature of science and making decisions using or about science. The report also builds upon the work of Project 2061 by defining science literacy as content specific knowledge and "understanding the nature of science" and the "role of science in society and personal life" (National Research Council, 1996, p. 2). The NRC's definition of science literacy included an understanding of the role of science in technology. Its definition also included the ability to read and understand news articles about science and to engage in scientific arguments. Nature of science skills were heavily emphasized requiring that students "explain and justify their understanding, argue from the data, and defend their conclusions, and critically challenge the scientific explanations of one another" (National Research Council, 1996, p. 50).

The NSES outlined by the NRC represents a comprehensive plan to address low science achievement as measured by the NAEP and TIMMS. The plan includes specific recommendations for curricular aims with an emphasis on science literacy, professional development, school-wide, and systematic reforms to improve science education. There are areas of concern with the NSES including the propensity of the report to rely solely on inquiry with minimal emphasis on reading, writing, and argumentation. Although critical to the nature of science curriculum, an overemphasis on inquiry can detract from science literacy because it can limit the time spent on the applicability of the content knowledge to current events.

A review of the literature by Roberts (2007) arranged the variety of definitions of science literacy as a continuum between Vision I and Vision II. Vision I is described as “literacy within science” that does not account for interaction of science with society (Roberts, 2007, p. 2) while Vision II is “literacy about science-related situations” (Roberts, 2007, pp. 2-3). The continuum approach differs from previous classification schemes by emphasizing the similarities between definitions rather than separation based on a matrix.

The aggregated definitions of science literacy produced by Project 2061 and the NSES fall within the framework of Vision I. Vision I is concerned with the development of scientific skill and understanding the structure of science (Roberts, 2007). The focus is the individual’s ability within science as a discipline while excluding the interaction between science and society. Vision I is associated with the cognitive deficit approach to science literacy in which knowledge exists independent of social interactions. This allows direct measure of science literacy with operational definitions that fall within the Vision I framework. An approach to science literacy that is limited solely to Vision I risks limiting the applicability of science to the life of a general citizen, thereby relegating the discipline to a series of facts.

Based on Vision II, science literacy is an understanding of the methodology or source of knowledge production that can be used to address problems in society, an idea that is commonly referred to as “science for specific social purposes.” Vision II focuses on individual decision-making, personal explanations, and everyday applications (Roberts, 2007). Vision II rejects the cognitive deficit model of science literacy because it considers knowledge situationally dependent. Longitudinal analysis of science literacy assessments that include components of Vision II found geographic variation with factor analysis demonstrating the relationship between curriculum emphasizing methodology and science literacy performance (Miller, 1998). Despite difficulty in measurement, Canadian educators have shifted the definition of science literacy toward Vision II with the inclusion of attitudes and real-world problem-solving skills in their K-12 learning standards (Roberts, 2007). A Vision II only framework limits the importance of science in science education.

Poor performance in science literacy has led to dire warnings of impending economic and social disaster. Miller (1998) compared science literacy to interest in politics suggesting that 80% of the public was inattentive and unconcerned with science policy resulting in low national performance. Concern for science literacy should focus on the 20% of the attentive public of which only 30% met the minimum standards for science literacy. The lack of clear goals and oscillation between student, teacher, and curricular expectations has created an untenable situation. Definitions of science literacy exist on a continuum between Vision I and Vision II depending on the degree to which science interacts with other disciplines, society, or social institutions. Although Vision I and Vision II represent extreme positions on the science literacy continuum, the inclusion of science inquiry, problem solving, and real-world decision-making within the NOS framework provides a measurable operational definition accessible to both visions (Roberts, 2007).

Science literacy remains a nebulous term that, although part of the educational lexicon for over sixty years, is still ill-defined. The classification systems of Pella (1966), Gable (1976), Laugksch (2000), and Roberts (2007) are useful in providing context and purpose behind operational definitions but have done little to standardize the approach to measuring science literacy. Lack of clarity in the term has led education researchers to "include everything possible in the definition" (DeBoer, 2000, p. 594). Lack of clarity in national performance expectations may be responsible for only one in five high school seniors reaching proficiency in science literacy (Grigg et al., 2006). The NAEP, PISA, TIMSS, and numerous other assessments have been developed to measure science literacy despite the lack of a common operational definition.

Measurement of Science Literacy

In order to measure science literacy, an operational definition for science literacy is required. Various instruments purport to measure science literacy, but they focus only on assessing content knowledge. Measurements of science literacy have taken multiple forms including surveys of the general public, content-dependent multiple-choice assessments distributed to K-12 students, or open-ended questions about the nature of science.

For the purposes of this study, science literacy is a construct that is independent of content knowledge to minimize variation based on course sequence. Science literacy is operationally defined in accordance with PISA's definition focusing heavily on NOS or the individual's ability to explain, evaluate, design experiments, and interpret data (OECD, 2017). NOS skills are a high level expectation. A national survey of adults in 1957 demonstrated that only nine percent of all adults understood an NOS approach to problem solving.

National and International Assessments

NAEP Science was developed by the U.S. Department of Education in 1969 to measure national progress in science for students in fourth, eighth, and twelfth grades every four years. A representative sample of students in each grade from across the country including every state and county is assessed since 1969. The National Assessment Governing Board is responsible for the development of the NAEP framework and assessment design. The NAEP framework guides the development of the assessment and remains consistent for a minimum of three administrations to allow for comparison between groups. The current iteration of the NAEP Science is divided into two domains: science content and science practices. Science content is divided further into physical science, life science, and earth and space science. The Science Practice section is divided into identifying science principles, using science principles, using science inquiry, and using technological design.

The NAEP is recognized as a valid and reliable assessment (Sireci, 1998) with a large amount of publicly available data for comparison to national and regional means. Science literacy was the goal developed by the AAAS and stemmed from student performance on the NAEP Science. The NAEP Science has been a useful assessment of science literacy for this study because of the immense amount of data and research associated with each administration that can be generalized to the larger population nationwide. The NAEP includes a question tool that allows the creation of customized or pre-selected assessments using the NAEP framework including questions that have been used in previous administrations of the NAEP.

Despite the advantages of using this valid, reliable, and nationally normed instrument, the NAEP Science was not appropriate for use in this study. The NAEP Science assesses content knowledge and lacks the independent measurement of scientific literacy necessary given the operational definition of science literacy used in this study. Questions in the NAEP mix content knowledge, NOS questions, and dual-coded questions that prevent the disaggregation of content knowledge from NOS skills. The NAEP science is an instrument designed to measure longitudinal changes in large populations over multiple years. The framework for the NAEP Science lacks the specificity in measuring science literacy independent of content knowledge failing to match the operational definition.

The Trends in International Mathematics and Science Study assessment has much in common with the NAEP. The TIMSS is an international assessment designed to measure longitudinal and large-scale changes in science and math performance. The TIMSS assessment is divided into a TIMSS Advanced and the TIMSS test. The TIMSS assessment is designated for use with fourth and eighth grade students making it an inappropriate selection for use with the participants in this study. The TIMSS Advanced is designed for high school students and is used to compare performance internationally. The framework for questions in the reasoning domain is consistent with aspects of science literacy including formulating questions, drawing conclusions, and designing investigations. The science reasoning domain is secondary to the TIMSS Advanced goal of measuring content acquisition and application in physics. It is estimated that less than 25% of the remaining participants would meet the requirement of physics course completion, thereby severely limiting the participant population. The inclusion of physics content does not match the operational definition of science literacy.

The Programme for International Student Assessment as developed by the Organisation for Economic Cooperation and Development purports to measure scientific literacy. The assessment is designed for 15-year-old students worldwide. Scaled scores are developed across content and process subscales that result in an overall science literacy score representing the average of the subscales. Content validity was established by having assessment items reviewed by educational representatives of all countries that use the assessment as well as a subject matter expert groups that independently verified question accuracy and alignment with the developmental framework (OECD, 2017). The items were then field tested to address potential bias with items not reaching the pre-established threshold removed from the 2015 administration of the PISA.

The PISA Science assessment has several advantages, which is why it was used in this study. The operational definition for science literacy in this study is taken from the PISA Science development framework. Although content knowledge is not removed from the PISA Science, the focus on science literacy as the aim reduces the impact of content knowledge on performance. The target audience for the assessment aligns with this studies' participant pool of high school science students. The overall science literacy score may have content dependency but the presence and reporting of data on subscales allows the use of the PISA. The PISA was the most appropriate large-scale assessment of science literacy for this study.

The PISA is unavailable in an intact form; there is variation in the number and types of questions used between countries, schools, and within a school. With the computer-based administration of PISA since 2015, the level of variation of the assessment has increased. This severely limits the use of the instrument for measuring changes in science literacy at the individual level. The purpose of the PISA is not to measure individual performance but rather to measure large-scale longitudinal changes in science education by country. Any variation of the PISA using a sample or selectively available questions would invalidate previously established validity and reliability.

National and international assessments such as the NAEP or PISA have well established procedures to ensure their validity and reliability in measurement of science literacy. These instruments also provide substantial data for comparison with international, national, and regional students. These assessments often incorporate dual-coded questions that assess both content knowledge and science literacy, confounding analysis of science literacy. These instruments are designed for longitudinal comparison and lack widespread availability.

Assessments Developed for Research

Improved performance on national and international assessments may be the ultimate measure of improved science literacy; however, the NAEP Science and PISA Science lack the responsiveness and contextual independence needed for this study. The longitudinal nature of national and international science literacy assessments requires the development of small-scale science literacy assessments. In recent years, there has been a large number of assessments developed to target specific aspects of the nature of science including: Nature of Science Literacy (NOSLiT), Science Literacy Concept Inventory (SLCI), the Test of Science Literacy Skills (TOSLS), and numerous others.

Nature of Science Literacy Test (NOSLiT). The Nature of Science Literacy Test (NOSLiT) is a 35-question selected response assessment developed by Wenning to measure high school students' scientific literacy. The instrument lacks an operational definition of science literacy and instead assesses nature of science skills as a proxy for science literacy. NOS represents a "central theme" of science literacy while "awaiting an operational definition of science literacy" (Wenning, 2006, p. 11). The instrument was piloted with 386 students from six high schools in central Illinois followed by a second pilot consisting of 354 students from the initial pilot. High school student performance ranged between 50-60%. The NOSLiT has been used as a measure of scientific literacy in high school courses at which an arbitrary 50% were used to determine if one was scientifically literate (*M* = 57.2%; Murti et al., 2018). Undergraduate chemistry students (*M* = 69.7%) were also assessed using the NOSLiT with no significant difference in results between genders or interest in science (Garner-O'Neale & Ogunola, 2015). A multiple regression function of predictor variables was performed to show that study habits, classification, gender, and interest in science did not have a significant impact on science literacy (p = .581).

The instrument was specifically developed for a target population that is similar to the one used in this study. Furthermore, nature of science is recognized as a key component of science literacy. The NOSLiT is freely available for and since it is an untimed assessment that takes approximately 30 minutes to complete, it fulfilled the need for an assessment that could be completed in one class period by this study's participants.

The lack of an operational definition that uses science literacy or a theoretical framework by which science literacy is measured limits potential use of the NOSLiT as an instrument. The instrument was developed by a single researcher, reviewed by a panel of undergraduate students, then piloted, which limits its content validity. The reliability of the instrument is also dubious because the reliability, after the first pilot of the instrument, was determined to be "an unacceptably low" number of 0.67, which led to substantial editing (Wenning, 2006, p. 23). A secondary instrument, Scientific Inquiry Literacy Test (SCInqLiT), was developed by Wenning to assess the skills of scientific inquiry within the larger framework of science literacy. This assessment had a high reliability (KR-20 0.88; Wenning, 2006) but it did not match the operational definition of science literacy used in this study

Science Literacy Concept Inventory (SLCI). The Science Literacy Concept Inventory (SLCI) is a 25-item selected response assessment developed by Nuhfer and colleagues to measure the science literacy of undergraduate students. The SLCI was developed based on 12 concepts that measure science literacy including nature of science, key terms and concepts, and interaction between science and technology. A panel of content experts representing a variety of disciplines developed 80 selected response items to be field tested. The questions were divided into two banks and a large-scale field test involving more than 20 colleges and universities was conducted. A 25-item assessment was developed in which items were mapped to specific concepts. The SLCI was capable of measuring within semester gains as shown with a pre-test (M = 71.70), post-test (M = 74.65) design with significant results, p < .01 (Nuhfer et al., 2016).

The SLCI was psychometrically validated as an instrument that measures science literacy. The SLCI was capable of measuring within semester gains as shown with a pre-test (M = 71.70), post-test (M = 74.65) design with significant results, p < .01 (Nuhfer et al., 2016). Reliability was determined to be acceptable at α = 0.86 (Nuhfer et al., 2016). Factor analysis showed that despite 12 separate concepts, the assessment loads onto a unitary construct.

The assessment correlated with factors that support higher levels of science literacy. A correlation between mean SLCI and mean ACT scores (r = 0.73-0.80) demonstrated a relationship between the selectivity of an institution and science literacy. Academic rank also showed significant differences with freshmen scoring the lowest and professors scoring the highest. The relationship between the number of science courses and SCLI score showed little variation until completion of a fourth science course. Various demographic factors were assessed with ethnicity, science commitment, a construct related to attitude, and English as a native language resulting in statistically significant differences in total SLCI score. A native

English speaker on average scored 7.2% higher when compared to non-native speakers and students pursuing a science major scored 5.26% higher when compared to non-science majors (Nuhfer et al., 2016, p. 151).

The SLCI has several advantages over the NAEP Science and PISA Science in that it demonstrates content validity and adequate reliability as a measure of science literacy. The instrument includes measures of NOS while incorporating scientific concepts and the interaction of science, technology, and society (Nuhfer et al., 2016). Differences established between groups are consistent with Gender, Ethnicity, English Language Proficiency, and Number of Science Courses Completed used in this study. The large sample size and number of institutions included in the development of the instrument support the generalizability of the correlations outlined by Nuhfer et al. (2016). The SLCI is an untimed assessment estimated to take approximately 30 minutes, which fulfills the set timeframe for participants in this study. The SLCI framework includes key scientific terms and concepts; it also incorporates content knowledge into a measure of science literacy.

The SLCI has factors that preclude its use in this study. The instrument was developed using undergraduate students, graduate students, and professors. The instrument has a high degree of difficulty with content area experts' (M = 90%) performance suggesting that the instrument may be inappropriate for use with a high school student. The development of the questions using multiple content experts resulted in the inclusion of questions requiring discipline specific knowledge that is unfamiliar to high school students would result in the introduction of a confounding variable. The SLCI framework includes key scientific terms and concepts and incorporates content knowledge into a measure of science literacy. The SLCI lacks widespread usage in the literature as an external validation of the properties assessed by developers of the instrument.

Test of science literacy skills. The test of science literacy (TOSLS) was developed by Gormally et al. (2012) to measure undergraduate biology students' science literacy. The TOSLS is a 28-question selected response assessment that includes measures of scientific and quantitative literacy while mitigating the influence of content knowledge. Science literacy is measured by nine subscales across two domains: process leading to scientific knowledge and organization, analysis, and interpretation of data. The operational definition for science literacy is based on that of the National Research Council (NRC), which includes "use evidence and data to evaluate the quality of science information and arguments" (National Research Council, 1996, p. 145). It also includes the National Assessment of Adult Literacy (NAAL) definition for quantitative literacy, which is "the knowledge and skills required to perform quantitative tasks" such as figuring a tip at a restaurant or the amount of interest on a loan (Kutner et al. 2007, p. 2). The inclusion of quantitative literacy is consistent with the expectation that scientifically literate individuals can organize and interpret data.

Construction of the TOSLS began with a national survey of undergraduate biology instructors (*n* = 188) actively participating in an undergraduate biology teaching association. A diverse sample of private/public, two/four year, and non-science/science majors responded to the survey, which consistently ranked nature of science as the most essential component of science literacy. Responses to the survey were used to construct the categories of skills in the TOSLS test, which were broken down into two domains that focused on common misconceptions surrounding science literacy at a post-secondary level. The first step including a wide range of faculty in identifying science literacy skills provides additional content validity to the assessment.

The TOSLS was initially piloted using 80 student participants with a selected response form and a constructed response form. Constructed response allowed follow-up interviews to refine items for a second pilot using a pretest/posttest design. A second, larger pilot was conducted in multiple undergraduate courses during Fall 2010, and Spring 2011 with the refined instrument. A pretest/posttest design for the Fall (n = 340) was conducted to examine the item discrimination index and quality of distractors for each selected response item. The process was repeated in Spring 2011 with students taking the pretest (n = 498) and posttest (n = 378). Following further revision, the TOSLS was reduced to a single version and administered as a pretest to undergraduate students in Summer 2011 (n = 70). A focus groups of students was used to adjust the wording and confirm that student reasoning matched the answer selection. A panel of experts in undergraduate biology education evaluated the TOSLS questions for accuracy, clarity, and fidelity to the assessment framework. The panel failed to reach 80% agreement for clarity with five items necessitating further revision. Student interviews were again conducted to ensure that students reached the correct answer using appropriate reasoning and problem-solving skills (Gormally et al., 2012).

The final version of the instrument was assessed using a pretest/posttest design demonstrating an average item difficulty of 0.59 (Gormally et al., 2012). Weighted means were calculated for the pretest (M = 60.72%), and posttest (M = 66.71%). A range of mean scores was established with significant variation by institution classification. The mean scores ranged from M = 42.50% for post-test scores at a mid-sized state college to M = 84.95% at a private research university (Gormally et al., 2012). An item discrimination index analysis was used to measure how well an item differentiated high and low scorers. Values below .2 indicated that the item did not sufficiently differentiate between high and low scorers (Ebel, 1965). The average item discrimination was .26 for the pretest and .27 for the posttest with a range from .05 to .41. Item discriminator indices were calculated for the TOSLS with only three questions--numbers 3, 15, and 24, falling under the .20 threshold. The TOSLS's validity in measuring student comprehension of understanding and ability to interpret basic statistics could not be established with 2 out of the 3 questions, failing to meet the established item discrimination expectations (Gormally et al., 2012).

An exploratory correlation was conducted using a simple three-question survey about students' science major, previous science coursework, and attitude toward science. Students previously enrolled in a science course scored significantly higher than those enrolled in a first undergraduate science course (p = .0254). Disaggregation of the data domain and subscale demonstrated significant differences between participants in their first science course and subsequent science courses for the NOS domain. The lack of statistically significant differences in quantitative reasoning between students who had completed an undergraduate science course and those enrolled in their first course may stemmed from the varied learning objectives of different undergraduate courses (Waldo, 2015).

The TOSLS was determined to be a reliable instrument sensitive to measuring within semester changes to science literacy. Reliability was determined using the Kuder-Richardson-20 formula for estimating internal reliability. Internal reliability is classified as good when $\alpha > 0.80$ (Cortina, 1993). The TOSLS was assessed to have acceptable reliability for both the pretest, α = 0.731, and posttest, α = 0.748, administrations. The failure of the instrument to reach $\alpha > 0.80$ indicates that the TOSLS may not measure a single construct. A follow-up exploratory factor analysis identified a single factor is likely the cause for the variability in scores (Gormally et al., 2012). The TOSLS is measures gains during a semester course by comparing pretest and posttest scores using a paired samples *t*-test. Students enrolled in a private research university did not demonstrate measurable gains, which may be caused by their high scores on the pretest (M = 84.63) and limited sample size (n = 50). This is consistent with the expectation of a measurable difference in science literacy during a college semester (Gormally et al., 2012).

The TOSLS is a valid instrument for measuring science literacy. Although originally developed for use in an undergraduate biology course, the instrument has been used in multiple contexts. However, the development of the questions using a single content area raises the possibility of bias because experts could have incorporated content knowledge into the science literacy assessment. During the development of a physics-specific assessment, Walsh, Quinn, Wieman, and Holmes (2019) assessed the TOSLS noting that the assessment requires evaluation of data, methods, and conclusions consistent with critical thinking skills (Walsh et al., 2019). Shortlidge and Brownell (2016), in their assessment of several instruments, noted that the TOSLS measures students' understanding of the process of science through science literacy, has been validated with multiple populations, is easy to score, and can be used in pre/post designs (Shortlidge & Brownell, 2016). Waldo (2015) administered the TOSLS at her institution to students in 18 natural science courses and validated the range of scores obtained during development. The study also demonstrated that students completing previous science coursework scored higher on the TOSLS assessment (Waldo, 2015). The consistency of the scores between students enrolled in similar universities provides validity for establishing comparisons across institutions. The TOSLS has also been used to assess undergraduate students' performance in science as an indirect measure of graduate teaching assistants' teaching competency (Graff, 2016).

The TOSLS has also been used as an outcome variable in measuring the impact of attitude toward science and science literacy. Student attitudes toward learning science was assessed with a single question asking students if the student liked to study science. Significant differences were reported for students that self-reported liking science (p = .003) when compared to students that reported not to like it very much (p = .028). Although indicative of the importance of attitude toward learning, a single question does not provide sufficient evidence linking attitude toward science and science literacy.

The TOSLS has been used as a measure of the effectiveness of curriculum changes. Auerbach and Schussler (2017) prepared a modified version of the TOSLS using 16 of the 28 items that directly corresponded with the objectives of the course. A comparison of scores between students (n = 156) enrolled in a traditional biology curriculum with students (n = 156) enrolled in a literacy-rich science curriculum demonstrated that both cohorts had significant increase in their science literacy scores between the pre/posttest assessments. The traditional curriculum participants resulted in a small average gain in student performance (M = 7%, d = 0.07) when compared to the literacy-rich participants' gains, (M = 13%, d = 0.71). There were no significant differences in scores based on students' gender, ethnicity, or educational classification (Auerbach & Schussler, 2017).

A sample of undergraduate non-biology major students (n = 164) was used to assess the connection between student course grades and their scores on the TOSLS. Students were told the TOSLS would either be an ungraded assignment (n = 83) or curved to 90% while weighing less than 3% of the final course grade (n = 81). The purpose was to measure the effect of student motivation on TOSLS performance to ensure department curricular decisions were made using valid TOSLS data. Data was collected from fourteen sections over a 2-year period with grading conditions applied at the course section level. At the end of the TOSLS administration, students were asked to rate their effort on a scale of 1-10. There were no significant differences between the ungraded assignment and graded assignment TOSLS scores (p = .416; Segarra et al., 2018).

The TOSLS was designed to assess undergraduate biology students' science literacy, but there is significant overlap between the domains and skills used to develop the TOSLS test and the framework used to develop the PISA Science and NAEP Science assessments. Nature of science, experimental design, interpretation of data, visualization of data, validity of sources, and communication of conclusions can be found in state and national standards (National Research Council, 2013; TEC, 2018). These skills are outlined in the Next Generation Science Standards as cross-cutting skills and in Texas as process skills. These skills also appear in the NAEP Science and PISA Science developmental frameworks (Neidorf et al., 2015; OECD, 2017).

The TOSLS has been used in a variety of undergraduate institutions (Coke, 2014; Segarra et al., 2018; Waldo, 2015) as well as in a high school setting (Chandler, 2017). Mean performance of science major students was consistent with high school (M = 61.59%) honors chemistry (n = 90) students in a pilot study (Chandler, 2017). The TOSLS was developed as a content neutral assessment of science literacy although questions that are framed in the context of biology. The participant pool for this study was limited to students completing a biology course to minimize the impact question context had on correctly answering the question.

The TOSLS was used as the outcome variable to assess the impact of a science literacy-based curriculum in lieu of content practice at the secondary level. Chandler (2017) administered the TOSLS and a content exam to honors chemistry high school students (n = 90).

Class period was the unit of randomization with 47 students in the literacy-based curriculum and 45 in the control curriculum (Chandler, 2017). Students' science literacy skills were assessed using the TOSLS and a modified version of the state assessment in chemistry at the end of the intervention period. Student scores were $M = 61.59$, $SE = 1.67$, in a manner consistent with undergraduate students. The literacy-based curriculum group ($M = 63.60\%$) scored higher when compared to the control group ($M = 59.39\%$), although the results did not approach significance ($p = .210$; Chandler, 2017).

Individual state standards, National Research Council, Next Generation State Standards, and Common Core State Standards all include learning objectives that directly relate to the skills outlined. The first domain is designed to measure a students' understanding of the Nature of Science, which is an understanding of the methods of inquiry to develop scientific knowledge and includes four subsequent categories of skills identifying a valid argument, evaluating sources, evaluating the use of information, and understanding research design. The nature of science skills is present in numerous state and national standards.

The Nature of Science subscale includes the expectation to "identify a valid scientific argument" (Gormally et al., 2012) that is consistent with the Next Generation Science standards (NGSS), which include learning objectives based on the assumption that "science knowledge is based upon logical and conceptual connections between evidence and expectations" and that "science disciplines share common rules of obtaining and evaluating empirical evidence" (National Research Council, 2012). Not all states have adopted NGSS, but several, including Texas, use definitions of science literacy that are similar to those used by NGSS. In the Texas Essential Knowledge and Skills for Science (2018) standards, science literacy is defined as "know[ing] that scientific hypotheses are tentative and testable statements that must be capable of being supported or not supported by observational evidence." The OECD has also distinguished between scientific arguments and key scientific competencies as the ability to "describe and appraise scientific investigations and propose ways of addressing questions scientifically" (OECD, 2007).

The TOSLS assesses students' ability to "evaluate the validity of sources" (Gormally et al., 2012). This skill is critical to the general population but is generally not included in high school science curricula. The NGSS includes links to key ideas in the Common Core State Standards including the ability to "delineate and evaluate argument[s] and specific claims in a text, including the validity of reasoning as well as relevance and sufficiency of evidence" (NGSS Lead States, 2013, p. 8).

The final test skill in the NOS domain requires students to "evaluate the use and misuse of scientific information" (Gormally et al., 2012, p. 366). In Texas, the curriculum requires students to "know that scientific hypotheses are tentative and testable statements that can be supported or not supported by observational evidence. As well as the ability to "draw inferences based on data related to promotional materials for products and services" (TEC, 2018, p. 2) within the Chemistry standards. The OECD also includes the ability to "use appropriate evidence for claims or conclusions" (OECD, 2007, p. 48). NGSS's definition of science literacy includes an understanding that science has limitations and that "science and technology may raise ethical issues for which science, by itself, does not provide answers and solutions" (NGSS Lead States, 2013, p. 127)

The second domain of the TOSLS focuses on students' ability to "organize, analyze, and interpret quantitative data and scientific information" (Gormally et al., 2012, p. 367). Students are expected to "create a graphical representation of data" (Gormally et al., 2012) that are consistent with the NGSS crosscutting concept of recognizing patterns for grades 6-12. There is substantial overlap between the TOSLS and NGSS regarding data analysis with the NGSS "graphs, chart, and images can be used to identify patterns in data" as well as "mathematical representations . . . needed to identify some patterns" (NGSS Lead States, 2013, p. 92). Those expectations are consistent with the curriculum in Texas, the largest non-NGSS state, with the Texas Essential Knowledge and Skills for Science (2018) requiring communication of valid conclusions through a variety of methods.

The connection between scientific literacy and quantitative literacy is evident in state standards, national standards, and the TOSLS assessment. Quantitative literacy skills are common in high school curricula with aims to help students develop the skill of "solv[ing] problems using quantitative skills, including probability and statistics" (Gormally et al., 2012, p. 367). The NGSS incorporates problem-solving skills into specific performance expectations including the ability to "use mathematical representations to support the claim that atoms, and therefore mass, are conserved during a chemical reaction" (NGSS Lead States, 2013, p. 91). These expectations are consistent with other state standards that require the students to express quantities and use mathematical procedures to perform calculations (Georgia Department of Education, 2016; Texas Essential Knowledge and Skills for Science, 2018)

The TOSLS is the appropriate tool for this study because it matches the operational definition for science literacy as outlined by the OECD. The TOSLS assesses students' ability to "justify inferences, predictions, and conclusions based on quantitative data" (Gormally et al., 2012, p. 367), which is also part of the scientific competencies of the OECD as the ability to "interpret data from related datasets presented in various formats" (OECD, 2007, p. 101). Ability to use quantitative data is also in the NGSS crosscutting concept for patterns that require students to identify patterns in evidence for cause and effect ("A Framework for K-12 Science Education," 2012). There is substantial overlap between the outcomes of the TOSLS and the process skills outlined in curriculum for Texas students to plan investigations, interpret data, and draw conclusions (Texas Essential Knowledge and Skills for Science, 2018).

Variation in the definition of science literacy, population assessed, and context of literacy questions has led to the development of multiple instruments to measure science literacy. Science literacy assessment has largely focused on the undergraduate population (Gormally et al., 2012; Impey et al., 2011, Nuhfer et al., 2016). Selection and use of an instrument depends on the reliance on content knowledge, participants' age, and the operational definition for science literacy on which the instrument is based.

Attitude Toward Science

Motivation and attitudes are non-cognitive factors that influence educational outcomes. Attitude and motivation are often used interchangeably in the common vernacular and early research. Early assessments combined both motivation and attitude in the same assessments such as Gardner's Attitude/Motivation Test Battery (Gardner, 1995). Motivation and attitude are divergent concepts with attitude defined as a generalized state of readiness related to a specific construct while motivation is goal directed behavior.

Motivation includes a combination of external and internal factors that initiate and sustain behavior that is directed toward a goal. Theories of motivation evolved from early research focusing on drive theory, to Skinnerian stimulus-response theory, to the current iteration emphasizing cognition as a mediating variable. Research on motivation began with studies of the reaction to external stimuli through observed behavior. This included research that would deprive "organisms living in an environment of limited resources" to examine the observable behaviors (Weiner, 1990, p. 617). Researchers focused on the need for water, food, and reaction to electric shock that would drive a behavior. External factors were expanded to include responses to learned stimuli. Behavioral psychology using the antecedent, behavior, consequence model of the behavior served as a precursor to expectancy-value theory accounting of internal mediators. Social Learning Theory was a response to criticism of pure stimulus response theory and the failure to account for complex behaviors such as language acquisition (Bandura, 1977).

Attitudes are internal judgements based on perceived control, perceived norms, and beliefs that form intentions to perform a behavior. Attitude is a cognitive mediator used to explain the apparent disconnect between antecedent and behavior (Ajzen, 2012). Early work in attitude focused on attitude as a single dimension representing all feelings about a topic. Further research demonstrated variation in attitudes emphasizing multiple factors that influence attitude including behavioral beliefs, normative beliefs, control beliefs, perceived behavioral control, subjective norms, and actual control.

Both motivation and attitudes include internal and external factors that influence behavior. Goal-directed behavior represents the primary difference between the two terms. Science literacy is rarely a function of explicit goal-directed behavior but measures of both motivation and attitude are affective variables that could influence science literacy.

Expectancy-Value Theory

Expectancy-value theory was developed by Fishbein (1975) as an extension of behaviorism in which pursuit of a task is mediated by the expectancy of success and the value of the task to the individual. Attitudes are developed as the sum of the subjective probability of the reinforcer combined with the value of the reinforcer. The expectancy-value theory was highly successful in correlating (r = 0.80) subjective probability and task value with direct measures of attitude (Ajzen, 2012) providing evidence that attitudes are formed based on expectancy-value theory. Expectancy-value theory was less successful in predicting behavior. In a meta-analysis of 184 data sets relating attitude to behavioral outcomes, a correlation of 0.27 indicated that there are additional variables that direct behavior (Greenwald, Poehlman, Ulmann, & Banaji, 2009).

In the 1970s, motivational theory began to shift away from external motivation to internal motivation with an emphasis on cognitive mediators (Dweck, 1986). Weiner (1990) stated that "this lack of separation, or confounding, between motivation and learning has vexed those interested in motivational processes in education, in part because learning is influenced by a multiplicity of factors" (Weiner, 1990, p. 618). The disconnect between attitude and behavior in expectancy-value theory "were mediated by higher mental processes" (Ajzen, 2012, p. 14). Support for expectancy-value declined with the studies showing that social learning provides vicarious reinforcement, the process in which learning occurs through observation and cognition instead of response to associated stimuli, impacting the likelihood that a behavior will be performed. Research on motivation and attitude shifted from observable behavior to exploring the cognitive mediators.

Self-Efficacy

Bandura suggested that a new construct of self-efficacy mediated efficacy expectations and outcome expectations (Bandura, 1977). Self-efficacy refers to the perception of one's ability to complete a designated task. Research had indicated that "actual competence does not strongly predict confidence of future attainment [of a skill];" instead, perception becomes reality (Dweck, 1986, p. 1043). By contrast, the perception of ability is strongly correlated with success, motivation, and achievement (Wigfield et al., 2009). The perceived difficulty of task can influence behaviors associated with that task. The "establishment, maintenance, and attainment of personally challenging and personally valued achievement goals" relates to intrinsic or extrinsic goals and can have a significant effect on cognitive performance (Dweck, 1986, p. 1040).

The difference in motivation for high and low achievers led to a shift in research that focused on the locus of control. Students with specific performance goals require greater sustained effort to complete a task, generally have lower persistence rates, sacrifice challenges for appearing successful, and struggle with confidence. Furthermore, their effort is negatively correlated with their own satisfaction. Performance goals also overwhelm intrinsic motivation. Performance goal motivation in schools leads to task avoidance and to a deficit of content knowledge and skills, thus leading to low achievement. A learning goal orientation has been shown to significantly impact the transfer and generalization of science content. Students who have a goal of increasing competence, not achieve a judgement, scored higher on generalized but untaught science content and demonstrated higher levels of persistence (Dweck, 1986). Achievement motivation is a highly complex model with many integrative parts including family, culture, perceptions, goals, task value, and previous performance. Discrepancies in expectancy-value theory along with cognitive mediators led researchers "to draw a distinction between two kinds of attitudes: general attitudes…and attitudes toward performing particular behaviors" (Ajzen, 2012, p. 15).

Theory of Reasoned Action and Planned Behavior (TRAPB)

The Theory of Reasoned Action (TRA), was developed by Fishbein and Ajzen (1975) to predict behavior. The TRA allows for the prediction of conscious behaviors by measuring attitude toward a construct through normative beliefs and motivation to comply with beliefs.

TRA predicts intentions within a variety of settings (Sheppard, Hartwick, & Warshaw, 1988). Two meta-analysis studies were conducted on the TRA model analyzing the relationship between intentions and behavior as well as attitudes and subjective norms with intentions. In a review of 87 studies, the correlation between intentions and behavior was 0.53 while the correlation between attitudes and subjective norms was 0.66 (Sheppard et al., 1988). The development of behavioral intentions is strongly correlated with attitude while behavior is only moderately correlated with intention. The concept of intention is “the immediate antecedent of behavior” (Ajzen, 2012, p. 18). The TRA model has a robust ability to predict behavior within the limitations of conscious behavior. Analysis of the model indicates the presence of a mediating factor that influences the development of intentions.

There is a gap between the intention to complete a behavior and the behavior; “internal factors, such as lack of sufficient willpower and perseverance or lack of requisite skills and resources as well as external factors such as failure to obtain needed cooperation from another person can interfere with planned behavior” (Ajzen, 1985). The Theory of Planned Behavior (TPB) expanded the TRA to include “perceived behavioral control as an additional predictor” (Ajzen, 2012, p. 17). The incorporation of perceived behavioral control is similar Bandura’s conception of self-efficacy in which the perception of one’s ability to accomplish a task directly relates to time on task and behavioral outcomes.

Including perceived behavioral control into the TRA enhances the predictability of the TRAPB model (Madden, Ellen, & Ajzen, 1992), uniting social learning theory with expectancy-value theory to form a robust theory of motivation. This model has been successful in predicting intentions and behavior in multiple behavioral domains (Ajzen, 2011). The TRAPB model includes behavioral beliefs, normative beliefs, and control beliefs that lead to intentions.

Research has been conducted in numerous fields regarding the predictive nature of the TRAPB model with a meta-analysis of multiple previous meta-analysis studies showing a "correlation of 0.53 between intention and behavior" (Ajzen, 2012, p. 19) with perceived self-control and ability as the moderators between the two.

Attitude about science content such as biology is difficult to extricate from attitude toward social issues dealing with science such as cloning. Beliefs can influence the empirical evidence that is used to justify a scientific conclusion (Sinatra et al., 2014). Measurements of attitude toward science should not include items in which the normative beliefs for morality could trump normative beliefs and intentions toward science. Science literacy is a highly complex construct in which skills and access to resources could limit correlation between intentions and behavior, thereby necessitating a multiple regression model.

Measures of Attitude Toward Science

Numerous instruments exist to measure student motivation to achieve and student attitude toward learning. This study limited the consideration of instruments to examine the MSLQ previously used in a pilot study (Chandler, 2017) and instruments that measure attitude towards science. The Attitudes Toward School Science Assessment (ATSSA), Behaviors Related Attitudes and Intention toward Science (BRAINS), Motivated Strategies Learning Questionnaire (MSLQ), and Simpson Troost Attitude Questionnaire-Revised (STAQ-R) specifically target attitudes toward science and are suitable for the participant pool age of 15-18. Early assessments of science attitudes differed widely in their construct assessing scientific attitudes, attitudes to science, attitudes to science careers, and attitudes to science instruction (Germann, 1988). These assessments measured the relationship between attitude and achievement using grades as a measure of achievement.

Attitude toward science in school assessment. The Attitude Toward Science in School Assessment (ATSSA) was developed to measure 7^{th}–10th graders' attitudes toward science as a single dimension. A panel of experts reviewed statements written by Germann revising the instrument to 24-Likert statements, a self-assessment of attitude, and a teacher assessment both from 1 to 10. A pilot of the ATSSA demonstrated that fourteen items loaded to general attitude toward science (Germann, 1988, p. 695). The 14-item instrument was field tested in four studies demonstrating a high reliability ($\alpha > 0.95$), loading to a single factor, with item total correlations of above 0.73. The ATSSA showed a high correlation with semester and lab investigation grades accounting for 16% of variance in classwork (Germann, 1988, p. 699).

The ATSSA represents a shift toward robust, psychometrically valid measures of attitude toward science. The loading of responses onto a single factor using confirmatory factor analysis provides evidence to support the assessment of a single construct. The instrument was developed with and piloted for students like those expected in this study. The high reliability and brevity of the instrument are also advantages.

The ATSSA's narrow definition of attitude toward science limits its usefulness in measuring science literacy. Also, in the ATSSA, the wording of statements like "Science makes me feel uncomfortable, restless, irritable, and impatient" (Germann, 1988, p. 701) can be confusing. These four descriptors represent different concepts today. Students with ADHD may report being restless and impatient with a positive attitude toward science. This ATSSA limits attitude toward science exclusively to the classroom, a limitation that is not supportive of the holistic nature of science literacy.

Behavior, related attitudes, and intentions towards science (BRAINS). The BRAINS assessment of science attitudes was developed in response to criticism of previous science attitude scales. Early measurement of science attitude combined cognitive construct of science attitude with observable behavioral outcomes. The focus on cognitive mediators of behavior required a shift in the theoretical framework used to develop measures of attitude toward science. BRAINS was developed with a theoretical construct centered on theories of reasoned action and planned behavior (TRAPB) model (Summers & Abd-El-Khalick, 2018). The TRAPB model suggests that background factors determine beliefs and those beliefs lead to attitude, perceived norms, and perceived behavioral control. These three factors determine the intentions of the individual, which is closely associated with behavior (Fishbein & Ajzen, 2010).

The BRAINS survey was developed for students of grades 5–10 as a cross-sectional instrument designed to provide longitudinal information while reducing complications associated with longitudinal studies. Attitude has been disentangled from associated observable behavior toward attitude as a cognitive mediator that cannot be directly observed. The shift in the theoretical framework used to guide the development of a science attitude assessment necessitates revision of existing instruments or the construction of new instruments.

BRAINS was developed by analyzing assessment items from seventeen existing science attitude assessments including the STAQ-R. Items that did not conform to modern standards for clarity in a Likert survey or that did not match the cognitive capabilities of students in elementary school were removed. A total of 74 items using a 5-point Likert scale were considered: 62 from original assessments, 16 were subsequently modified from the original, and 12 items were written by the researchers. Items that matched to constructs in the TRAPB model were selected from the original instruments and included normative beliefs, control beliefs, behavioral beliefs, perceived behavioral control, perceived norms, attitudes toward behavior, and intentions. These items were subjected to screening for modern standards of clarity and to match the cognitive capabilities of students in elementary school (Summers & Abd-El-Khalick, 2018).

An expert panel consisting of researchers in science education, science attitudes, and science college faculty members reviewed the instruments for face validity. The panel's revision of the items resulted in a total of 59 items remaining in the assessment pool. An initial pilot was conducted (*N* = 151) using two sections of 3rd graders (*n* = 45) and five sections of middle school students (*n* = 106). A sample of students was selected and interviewed regarding a subset of 20 items to modify and clarify individual items (Summers & Abd-El-Khalick, 2018).

A second pilot study was conducted using students from 5th to 10th grades from geographically diverse areas of Illinois. Students (*N* = 1292) from 68 sections participated in the second pilot of the BRAINS to assess psychometric validity. A maximum likelihood robust estimator confirmatory factor analysis (CFA) was conducted to assess the content validity of the instrument without the assumption of normality (Li, 2016). CFA suggested that an overall construct was present, but the instrument failed to load onto a single construct. Items that did not load or were duplicated were removed from the assessment. The five-factor loading for theoretical constructs was reported in ranges that exceeded 0.4, which indicated a five-factor load onto a single unitary construct. This demonstrates the BRAINS is internally consistent and unidimensional (Atwater, Wiggins, & Gardner, 1995). Finally, items that loaded onto multiple factors were removed, resulting in a 30-item final instrument (Summers & Abd-El-Khalick, 2018). The final instrument was assessed to fit the model outlined by TRAPB with Root Mean Square Error of Approximate (RMSEA) with values below 0.05 representing a good fit. The model indicated BRAINS was a good fit for the TRAPB, RMSEA = 0.04. A comparative fit index (CFI) test was conducted with a value of 0.95 demonstrating good fit with the model.

Internal reliability was assessed by subscale with estimates ranging from α = 0.7–0.91 (Summers & Abd-El-Khalick, 2018).

BRAINS was constructed using modern interpretations of attitudes based on a cognitive approach to attitude. The BRAINS assessment is a 30-item Likert survey designed to provide longitudinal information on a wide-ranging population with a single administration. The assessment has face validity with experts, is a good fit to a robust model, and demonstrates strong reliability and validity. The target population are students from 5th to 10th grades and is appropriate for use with higher grades. The instrument was developed by incorporating previous assessments on attitude toward science, matching the assessment to a theoretical framework, and establishing modern psychometric validity. Despite the advantages of the BRAINS, its use in the literature is limited (Summers & Abd-El-Khalick, 2018).

Motivated strategies learning questionnaire (MSLQ). The MSLQ was developed in 1986 to assess the motivation and learning strategies that college students use (Pintrich & DeGroot, 1990). It consists of 81 questions divided into 15 scales in two major categories. All questions consist of a 7-point Likert scale ranging from *not true at all* to *very true*. In the MSLQ, definitions of motivation and learning strategies are based on social cognitive theory. There are six motivation subscale measures: intrinsic goal orientation, extrinsic goal orientation, task value, control of learning beliefs, self-efficacy for learning and performance, and test anxiety. The motivational subscales are based on expectancy, value, and affect. The MSLQ has three general types of subscales -- cognitive, metacognitive, and resource management. These subscales are further divided into rehearsal, elaboration, organization, critical thinking, metacognitive self-regulation, time and study environment managements, effort regulation, peer learning, and help seeking.

The MSLQ was developed as a social cognitive tool that considered student interaction with information as a dynamic interaction (Duncan & McKeachie, 2005) and has been correlated with motivation for biology (Solmaz, 2015). It considers a course the highest level of analysis and is not meant to be generalized beyond a course. The MSLQ has many different uses and has been used in a variety of studies. Generally, the MSLQ is used to measure the effect of different instructional designs on student motivation.

In 1990, using 356 students, researchers assessed the internal consistency, reliability and predictive validity of the MSLQ across multiple domains (Pintrich, Smith, Garcia, & McKeachie, 1993). Confirmatory factor analysis was used to assess the motivational and learning strategies models for validity. The internal reliability of the MSLQ differed across domains with the motivation subscale demonstrating a greater internal reliability than the learning strategies domain. Learning strategies had lower internal consistency due to variability in the help seeking subscale (Pintrich et al., 1993). Predictive validity indicated a weak relationship between most of the subscales and final course grades. Self-efficacy for learning and performance had the highest correlation with final course grade in the motivation scale (r = 0.41) indicating a moderate positive relationship. The highest correlated values on the learning strategies scale was a weak relationship between effort regulation, metacognitive self-regulation, and final course grades. The correlation of the subscales for each theoretical construct indicated that MSLQ was a valid measurement of motivation and learning strategies (Pintrich et al., 1993).

In an independent validity verification study for the Turkish version of the MSLQ, İlker, Arslan, and Demirhan, (2014) used the MSLQ with 1605 high school students. They used a confirmatory factor analysis to compare the scores of Turkish students to the general model put forth by Pintrich et al. (1993) to ensure validity. Multiple tests of the model, including a comparative fit index test, confirmed the validity of the MSLQ model. Additionally, Cronbach alpha coefficients were calculated for the scale, dimensions, and subscales indicating a high level of internal reliability (İlker, Arslan & Demirhan, 2014).

The MSLQ has well established internal reliability with coefficient alpha values ranging from 0.52 on the low end to 0.93 on the high end (Pintrich et al., 1993). Only help-seeking had poor internal reliability while self-efficacy had excellent internal reliability. The reliability of the MSLQ has been tested across multiple languages in secondary institutions as well as universities. Ilker et al. (2014) found that the Turkish version of the MSLQ had a high reliability (α = 0.70) for all dimensions (Ilker et al., 2014).

Crede and Phillips (2011) performed a meta-analytic review of the MSLQ in 2011 reviewing the findings of 67 studies and 2158 correlations of nearly 20,000 college students. The results of the meta-analysis indicated that the relationship between the MSLQ and academic performance were moderate at best (Crede' & Phillips, 2011). Metacognitive strategies, time spent studying, study environment, and effort had the most predictive validity in academic performance. The MSLQ provides important information regarding students' learning strategies and motivation but that information is generalized rather than content specific. Because of its predictive validity, face validity, and high reliability, the MSLQ is used widely. The instrument subscales of expectancy, value, and affect are consistent with those outlined in the TRAPB model.

The MSLQ total measures motivation and learning strategies by using subscales to derive information regarding attitude. Although motivation and attitude are similar constructs, a measure of motivation toward science is inconsistent with the theoretical framework outlined by the TRAPB model. Studies indicate that motivation and attitude are domain specific constructs

(Green, Martin, & Marsh, 2007). The MSLQ is not a discipline specific measurement of motivation toward science but instead applies general principles of motivation to a science class. The MSLQ contains 71items, which, while thorough, exceed the time limitations of this study.

Simpson-Troost attitude questionnaire (STAQ). The STAQ was developed by Simpson and Troost as part of a multiyear longitudinal study to identify and predict the factors that influence science attitudes and motivation toward science achievement among 6–10th grade students. The STAQ was developed using information from a single school district in North Carolina that was selected from many schools that includes students of geographic, ethnic, and economic diversity. A total of twelve different questionnaires were developed using a sample of approximately 4500 participants. The final version of the STAQ contained four domains: science, family, self, and school with a total of 60 selected response items. The internal consistency of the attitudes survey was above 0.90 for many subscales peaking at 0.94 for the science affect scale (Simpson & Troost, 1982). The STAQ used a reliability cut score of $\alpha = 0.35$, which is well below modern acceptable values of $\alpha = 0.70$ (Owen et al., 2008). The STAQ was developed during a time in which attitude was closely linked with behavioral outcomes as evident in the subscale's achievement motivation, anxiety, and response to external stimuli, which are measured with behavioral components (Simpson & Troost, 1982, p. 772).

In a psychometric reappraisal of the STAQ, Owen and colleagues administered the STAQ to middle school students (n = 1812) from a wide range of schools and districts. The STAQ was modified from the original 59 questions to 57 by removing a duplicated item, and the removal of a potentially objectionable item at the request of one of the participating school districts. A confirmatory factor analysis of the original 14 factor model was conducted with all data, including students, not responding to all questions, which resulted in an unsupported model

(χ^2 = 4,986.8, *df* = 1,234, $p < .001$). Exploratory analysis was conducted using half of the original data sample by reducing the number of factors by one to determine the number of factors that should be included in a model. The reevaluation of the instrument used a cut score of 0.70 (Owen et al., 2008 p. 1081). A 10-factor solution determined by exploratory factor analysis was reduced to a 5-factor model which was then tested using confirmatory factor analysis against the second half of the data set (χ^2 (203) = 853.9, $p < .001$, CFI = 0.94 RMSEA = 0.043). The resulting instrument, Simpson Troost Attitude Questionnaire Revised (STAQ-R), has been used as an updated measure of science attitudes

The STAQ was administered to students (n = 186) including 9th and 10th grade students to assess the reliability of the subscale findings using a test/re-test design. Students were administered the STAQ twice with one-week duration between administrations. The reliability was calculated using both Cronbach's alpha and Hancock and Mueller's structural equation modeling H. All five of the revised subscales demonstrated adequate reliability and stability using both Alpha and H assessments (α = .70- .90).

A modified version of the STAQ-R, adding 8 cultural and ethnic questions unique to Iran, was used to measure motivation of Iranian biology students. An initial pilot of the instrument demonstrated internal reliability similar that of the original instrument (α = 0.85). The instrument was administered to 185 12th grade students between the ages of 17 and 18 and compared against their achievement in biology. Regression analysis showed no significant correlation between student performance on the STAQ-R and achievement in their biology class (r =.12, $p < .5$). Disaggregation of the subscale data indicated that the *biology is fun* subscale was significantly and positively correlated with achievement in biology (r = 0.304, $p < .05$; Nasr & Soltani, 2011). This study demonstrated that the STAQ-R is an appropriate instrument for measuring science motivation with students up to 18 years of age. However, the study failed to clearly define the dependent variable of science achievement. Science achievement could be performance on a standardized assessment, course grade, or some other unaccounted-for factor. The lack of a clear definition limits the applicability of the Iranian study to this study.

In a study of urban elementary students ($N = 63$), gains in the STAQ-R were used to develop a model that would predict the acquisition of content knowledge. The STAQ-R accounted for 19.1% of the overall content knowledge acquisition. The self-directed effort and peer model subscales of the STAQ-R were shown to account for the greatest variance in content knowledge acquisition with self-directed effort accounting for 11.9% of the variance (Newell et al., 2015). The analysis also indicated that *science is fun* may be a suppressor variable because of "medium sized beta weights and close to zero squared structure coefficients" (Newell et al., 2015, p. 220). This study demonstrated that the STAQ-R is a sufficiently robust measure of semester gains in motivation for students as young as those in elementary school. Gains in the STAQ-R, although different from science achievement, may align with science literacy in a similar manner.

Science Literacy Model Factors

Limited research has been conducted to explore variations in science literacy outside of national and international assessments. Although substantially different in outcome, science literacy has been confounded with science achievement. The literature was reviewed for variables commonly collected by schools for differences in science literacy or science achievement when science literacy information was unavailable. Though the differences between science achievement and science literacy are significant, there is an overlap in performance skills such as evaluating a hypothesis or interpreting data.

Gender

National and international assessments consistently demonstrate significant differences between males and females in science achievement (Beaton & Chromy, 2010; Campbell et al., 1997; OECD, 2016; U.S. Department of Education, 2012). In a recent administration PISA Science and NAEP Science, males performed seven points higher on the PISA Science and five points higher on the NAEP Science than females (Kuna et al., 2018; OECD, 2016). The TIMSS has also shown significant differences in science achievement by gender with males outperforming females in a statistically significant manner (Beaton et al., 1998).

Gender differences remain consistent regardless of participant age (Campbell et al., 1997). A review of existing records from the Early Childhood Longitudinal Study demonstrated significant gender differences in science content knowledge using a large (N = 8,741) national sample. The weighted averaged of male students (M = 59) was significantly higher when compared to female students (M = 56) although the effect size was small (η^2 = 0.006; Kohlhaas, Lin, & Chu, 2010). Tong and colleagues found gender differences in science achievement as measured by district benchmarks and the TAKS science test as early as 5th grade with males outperforming females (Tong et al., 2014). In an assessment of science literacy for secondary science teachers, Coke (2014) found that male teachers (M = 88.1) outperformed female teachers (M = 82.5) on the TOSLS).

A cross-national study of Anglo countries determined statistically significant differences between men and women (Hayes & Tariq, 2001). There are item-specific differences between men and women, with females performing higher on biology content questions (Hayes & Tariq, 2001). In an ongoing twenty-seven-year analysis of science literacy trends among undergraduate students, Impey et al. (2017) demonstrated that men outperform women by approximately 5% in science literacy assessments, but those findings were not statistically significant after controlling for other factors

There are discrepancies in the general trend of significant higher performance on science literacy and achievement for male students. The STAAR Biology assessment used to assess content knowledge deviated from the established trend. Female students had a mean scaled score 80 points higher than male counterparts (Texas Education Agency, 2019b), which was consistent with findings indicating that gender discrepancy is largely context specific (NAEP, 2015a). The SCLI also found no statistically significant difference between male (M = 68.5) and female performance (M = 68.2; Nuhfer et al., 2016). The differences are content-dependent with physical science difference representing the largest difference in performance (OECD, 2016). Possible explanations for these differences include willingness to make blind guesses on selected response items (Mondak & Ganache, 2004) or differing attitudes toward science (J. Osborne et al., 2003).

Gender differences can also account for variation in the effect of subsequent predictor variables. Among male students, there was a moderate correlation between interest and achievement for chemistry and physics while there was no significant correlation among female students (Larson et al., 2014). Gender discrepancies in science achievement and science literacy assessments have persisted over time across multiple measures. These disparities necessitate the inclusion of gender into any model meant to predict science literacy of secondary students. The identification of gender as a factor in a model of science literacy warrants a reconceptualization of the approach to science teaching and provides additional support to attempts to promote female participation in science careers.

Ethnicity

National and international assessments consistently demonstrate significant differences in science achievement among ethnic groups (Campbell et al., 1997; OECD, 2016). The most recent results of the NAEP science continue to show significant differences between groups at the high school level with Asian students scoring an average of 39 points higher when compared to Black students (Kena, et al., 2016). These findings are consistent with the administration of the 2015 PISA in which Asians achieved the highest mean score followed by White students, Hispanic students, Multiracial students, and Black students (OECD, 2016). This trend occurs at lower grades with only 37% of African American students and 48% of Hispanic students obtaining proficiency on the NAEP Science 8th grade assessment compared to 80% of White students doing so (U.S. Department of Education, 2012) The achievement gap on these assessments has narrowed over time (Campbell et al., 1997), but that trend has stalled in recent administrations (Kena et al., 2016). Using the Early Childhood Longitudinal Study, Kindergarten Class of 1998-99, Kohlhaas et al. (2010) demonstrated significant variance between ethnicities as early as fifth grade using Science Item Response Theory with the largest difference found between White (M = 63, SD = 11.42) and African American (M = 47, SD = 13.81) students with a medium-large effect size of η^2 = 0.0116.

Results of the recent STAAR Biology assessment, which is used to assess content knowledge, mirrored national and international trends. Asian students had the highest mean scale score that was 486 points higher than Native Hawaiian or Pacific Islander students (Texas Education Agency, 2019b). The SLCI demonstrated significant differences in science literacy when disaggregated by ethnicity ranging from 52.81 for Middle Eastern students to 71.27 for Caucasian students (Nuhfer et.al., 2016). Differences between ethnicities is largely attributed to confounding factors including language acquisition, school quality, and economic and social factors (Peng, 1995). Attempts to control for these factors have reduced but not eliminated differences between groups (Nuhfer et al., 2016).

Differences in performance based on ethnicity are inconsistent based on the design on the study. A study measuring the impact of curriculum change on science literacy reported no significant differences between white and non-white students as measured by changes in the TOSLS over a semester (Auerbach & Schussler, 2017). The study did not disaggregate overall performance by ethnicity due to the focus on the cohort and curricular change. Differences in science literacy and achievement remain, which warrants their inclusion in a model of science literacy and is consistent with the models used to predict literacy (Mohadjer et al., 2009). The identification of ethnic differences in a model of science literacy would support further exploration of sociocultural factors correlated with ethnicity and would support targeted intervention programs.

English Proficiency

The impact of English Proficiency on science literacy has not been well documented; however, content area literacy depends on students' ability to access information through reading, writing, and argumentation, which places English Language Learners (ELL) at a disadvantage on science literacy assessments (Adams & Pegg, 2012). Over 20% of students scoring below Basic Science Proficiency on the NAEP Science reported speaking a language other than English at home. This same group of students also had reduced likelihood of scoring at a level of *proficient* or above (Grigg et al., 2006).

In a comparison of SLCI performance between different ethnic groups, Nuhfer et al. (2016) measured the relationship between ethnicity, first generation college student status, and student's English language proficiency. The largest demographic factor that affected students' SLCI development was being a native speaker of English, which provided a 7.2% advantage over non-native speakers (Nuhfer et al., 2016).

During the 2014-2015 school years, the number of students classified as ELL rose to 9.4% (Kuna et al., 2018). The percentage of ELL students in Texas has risen to 15.5% (Kuna et al., 2018). Nationwide, students of grades 10–12 represent between 4.1–6.0% of the total participant population with 76.6% speaking Spanish (McFarland et al., 2017). Data from the most recent administration of the STAAR Biology shows students classified as Limited English Proficiency have a mean scale score that is 166 points lower than those of non-LEP students (Texas Education Agency, 2019b). The classification of LEP is reviewed annually with students progressing to *monitored* status with all students of *monitored* status outperforming the state mean scale score (Texas Education Agency, 2019b).

The differences between students classified as LEP and native speakers can result in a distinct advantage for native speakers who can understand nuance, cultural context, and academic vocabulary. These factors are consistent with NAAL definition of literacy focusing on communication in society and warrant their inclusion in an English proficiency model used to predict science literacy. Identification of English proficiency as the predominant factor in science literacy would result in a shift in approaches to science and language interventions.

Economic Need

The number of students who qualify for economic assistance has dramatically increased with 74.3% of school lunches served being subsidized through the Free and Reduced Price (FARP) program. Students who are eligible for participation in FARP are classified as an individual with economic need. Students are eligible for participation in the FARP based on household income that is equal to or lower than 185% of the federal poverty line.

Students participating in FARP had an average scaled score on the PISA Science that was over 100 points lower than the score of students not enrolled in FARP (OECD, 2016). In an examination of TIMSS data, Beaton et al., (1998) found that students reporting "more educational resources in the home had higher mathematics and science achievement" when compared to students with limited resources (p. 71). The variation in science achievement, as measured by the PISA 2015, indicated nearly 1.5 standard deviations in score between disadvantaged school and advantaged school when comparing socioeconomic profiles (OECD, 2016). Kohlhaas et al. (2010) found significant variance between students at or above the poverty line ($M = 61$, $SD = 12.77$) and those below the poverty line ($M = 47$, $SD = 14.79$) with a moderate effect size ($\eta^2 = 0.056$).

In a review of OECD data from PISA administration for Nordic countries, the researchers disaggregated the commonly used terms related to socioeconomic status to demonstrate the effects of cultural capital, social capital, and economic capital as separate entities. Economic capital was responsible for a small amount of the variation in PISA science scores ($R^2 = 0.02$– 0.10) for Nordic countries (Turmo, 2004).

Analysis of the 2012 Texas Assessment of Knowledge and Skills showed that economically disadvantaged students ($M = 41.1$) are significantly more likely to be classified as "not college ready" when compared to their peers ($M = 22.1$, $p < .01$; Lee & Slate, 2014). Fifty-seven percent of students completing the STAAR Biology assessment in 2018-2019 were classified as being in economic need (Texas Education Agency, 2019b). Data from the most recent administration of the STAAR Biology shows students classified as Economically

Disadvantaged have a mean scale score that is 168 points lower than those of non-disadvantaged students (Texas Education Agency, 2019b).

Economic disparity between student populations represents a widespread achievement gap in multiple subjects (Kolhaas et al., 2010). Students with economic need are provided with less instructional minutes (Blank, 2013) and have less opportunity to learn, which has accounts for nearly one-third of the literacy discrepancy (Schmidt, Burroughs, Zoido, & Houang, 2015). Identifying economic status as a variable impacting science literacy may necessitate intervention outside of the traditional school day to provide meals, resources, housing, and other interventions that address poverty more directly.

Science Content Knowledge

The TOSLS is designed to assess science literacy independent of content; however, it the questions are formulated using biology as the context. Biology Texas Essential Knowledge and Skills (TEKS) “analyze, evaluate, make inferences, and predict trends from data” (Texas Essential Knowledge and Skills for Science, 2018), which is very similar to the categories of science literacy outlined by Gormally (2012) in the development of the TOSLS. Gormally et al. (2012) defined science literacy as the ability to “justify inferences, predictions, and conclusions based on quantitative data” (p. 367). The overlap in definitions suggests that performance on the STAAR Biology exam may correlate with TOSLS performance but the dual role of content and process skills in Biology STAAR must be considered. The Biology STAAR blueprint (Texas Education Agency, 2018) used to develop each iteration of the assessment indicates that questions are linked to specific content standards with process skills “incorporated into at least 40% of the test questions” (Texas Essential Knowledge and Skills for Science, 2018). The

STAAR Biology assessment is also a measure of science content knowledge as well as science process and skills.

A multiple regression study to determine the influence of predictive variables on science achievement assessments demonstrated that previous science performance was highly predictive (R^2 = .27 - .39) of future performance accounting of the variation in performance (Larson, Stephen, Bonitz, & Wu, 2014). A correlation between mean performance on the SLCI and ACT composite scores was conducted by Nuhfer et al. (2016) at the institutional level. ACT scores were correlated with the SLCI and demonstrated ACT composite scores (r = 0.79, p < .01) that contained Science and English Language Arts had a higher correlation coefficient than the SAT Math scores (r = 0.69, p < .01).

Question context, including vocabulary and content, may influence individual performance on science literacy assessments. Biology content may act as a hidden curriculum that unduly influences science literacy assessment as measured by the TOSLS. Identifying science background knowledge as a variable impacting science literacy may support the development of science literacy by emphasizing the acquisition of content knowledge as a framework for science literacy thereby necessitating increased instructional time.

English Language Arts Content Knowledge

The National Assessment of Adult Literacy (NAAL) definition of literacy is to “use printed and written information to function in society, to achieve one's goals, and to develop one's knowledge and potential” (Kutner et al., 2007, p. 2). This represents a broad view of literacy as applied to generalized knowledge. Science literacy is the discipline-specific application of these principles to acquire, interpret, and communicate scientific information and ideas. Performance on the State of Texas Assessment of Academic Readiness English Language

Arts assessment (Texas Education Agency, 2019c) was included in the model of variables in this study as a standardized measure of students' ability to understand, analyze, and compose informational texts. The learning standards for the STAAR ELA include the ability to "synthesize and make logical connections between ideas and details in several texts selected to reflect a range of viewpoints on the same topic and support those findings with textual evidence" consistent with evidence-based reasoning outlined in science literacy assessments (Texas Essential Knowledge and Skills for English Language Arts and Reading, 2010).

Tong et al. (2014) evaluated the impact of a literacy-integrated science curriculum with middle school students by reading expository texts, discussion, and writing. The intervention was assessed using district benchmarks, Dynamic Indicators of Basic Literacy Skills (DIBLS), and Texas Assessment of Knowledge and Skills (TAKS), which was a precursor to the current iteration of STAAR. A regression model showed that students participating in the intervention were more likely to pass district benchmarks and more likely to pass the science TAKS test although the statistical significance was marginal (Tong et al., 2014).

Analysis of NAEP results and methodology outline substantial bias of the instrument against English language learners. Researchers determined that incorporation of informational literacy concepts into a science writing course improved student confidence and science literacy skills, which were measured through portfolios (Klucevsek & Brungard, 2016). The use of informational literacy as a proxy for science literacy is consistent with the understanding of science literacy as a multimodal construct (Holbrook & Rannikmae, 2007). Content area literacy requires proficiency of language, and science is no exception. Nature of science cannot be assessed without "talking, writing, and reading" (Yore & Treagust, 2006).

Science literacy as a multimodal construct that requires the incorporation of multiple literacies (Holbrook & Rannikmae, 2007). Science achievement state assessment scores exhibited a stronger correlation with reading skills than previous science knowledge assessments (O'Reilly & McNamara, 2007). Reading is a source of scientific knowledge while writing is a metacognitive process that enhances understanding. Identifying generalized literacy skills as a variable impacting science literacy may support the development of a horizontal curriculum in which content acquisition is viewed to develop generalized and science-specific literacy skills.

Algebra Content Knowledge

Performance on the State of Texas Assessment of Academic Readiness Algebra 1 assessment (STAAR Algebra 1) was included in the model of variables of this study (Texas Education Agency, 2019a). Science literacy, as a measured by the TOSLS, includes quantitative literacy (Gormally et al., 2012). The learning standards for the STAAR Algebra 1 include "relations represented verbally, tabularly, graphically, and symbolically," which are consistent with the interpretation of data necessary on science literacy assessments (Texas Essential Knowledge and Skills for Mathematics, 2015).

The design of the NAEP precludes simple correlation between the NAEP Math and Science scores due to the variation in participants. Although given on the same schedule, students completing the NAEP Science are not the same students completing the NAEP Math. Analysis of national trends show that performance in mathematics from 1978 to present ranged in scaled scores from 298 to 308 with an overall rise in scores during that time frame (U.S. Department of Education, 2019). The general improvement in mathematics performance is consistent with the slight increase in science performance.

In an analysis of existing TIMSS data, Wang and Ma (2016) correlated math performance and science performance for 7th grade students from 35 countries. This study correlated plausible score combinations from multiple subjects into single variables demonstrating a strong correlation between math and science performance $\beta = 0.791$, $p < .01$ (Wang & Ma, 2016). The relationship between math achievement and science achievement depends on disciplinary content. A correlational study found that undergraduate performance on a math assessment moderately correlated with student science achievement gains on the Force Concept Inventory assessment (Hake, 1998). In a quasi-experimental study of nearly 1200 8th grade students, the researchers found significant improvement in mathematics scores when the science curriculum was infused with math concepts (Burghardt, Lauckhardt, Kennedy, Hecht, & McHugh, 2015). The study found the greatest gains in high-level reasoning domains with the greatest improvement shown by the lowest students initially in the lowest quartile on the pre-assessment.

Science and math have been historically linked with significant overlap in curriculum. A Mathematical Models course specifically outlines this link: “The student applies mathematical processes with algebraic techniques to study patterns and analyze data as it applies to science” (Texas Essential Knowledge and Skills for Science, 2015). Elements from the math curriculum can be found in five of the nine elements the TOSLS assesses as a measure of science literacy. Identifying mathematics achievement as a variable impacting science literacy may support the extension of successful integrated math/science curriculum to high schools (Czerniak, Weber, Sandmann, & Ahern, 1999).

Grade Point Average

Grade Point Average, GPA, is calculated by converting numerical grades from a 0–100 scale to a 4.0 scale: 90– 100 receives 4.0 points, 80–89 receives 3.0 points, 70–79 receives 2.0

points, and grades below 70 receive 0 points. High school GPA is a combination of cognitive and non-cognitive factors, including consciousness (Noftle & Robins, 2007) that represent one measure of achievement. High school GPA has correlated well with successful transition to college in predicting freshmen GPA ranges (Noble & Sawyer, 2002).

Impey et al. (2011) assessed the relationship between undergraduate GPA and science literacy finding no statistically significant relationship between the two. The definition used by Impey et al. to assess science literacy relied heavily on the use of content knowledge diverging from the operational definition used in this study. The use of self-reported GPA had the potential to introduce a confounding variable in analysis as self-reported GPA may be under or over reported by students of different demographics (Caskie, Sutton, & Eckhardt, 2014)

GPA can be overemphasized; however, it represents a combination of cognitive and non-cognitive factors like science literacy. GPA is an imperfect measure of academic achievement that includes non-cognitive factors such as attitude, motivation, and support at home. The SLCI was shown to correlate with ACT Composite scores ($r = 0.79 – 0.85$, $p < .01$) linking science literacy and academic achievement (Nuhfer et al., 2016). Identifying GPA as a factor impacting science literacy would provide evidence supporting the impact of affective variables supporting the addition of science appreciation in science curriculum.

Number of Science Courses Completed

The percentage of high school students completing science courses in high school has dramatically increased between 1990 and 2009. The percentage of students completing biology courses increased from 91% to 96%, percentage of those taking completing chemistry courses increased from 49% to 70%, and the number completing physics courses increased from 21% to 36%. Despite these increases in the number of students taking additional science courses during this period, PISA results from 2000, the earliest available, as well as from 2009, have shown no improvement in students' science literacy (OECD, 2000; OECD, 2016).

In a study of student teachers in the United Kingdom, (Murphy, Beggs, Hickey, O'Meara, & Sweeney, 2001 compared the scores on a science assessment with a "low level of difficulty" between a group that had compulsory science education between ages 6-11 and a group that did not have compulsory science education (p. 194). Murphy et al. found that students with compulsory science education scored significantly better (M = 66%) than the group that did not have compulsory science education (M= 60%), which provided evidence of long-term content retention resulting from additional science coursework. Secondary analysis of NAEP Science data demonstrated a decrease of 2.3 hours of instructional time in elementary science that correlated with a 12-point drop in science achievement (Blank, 2013).

The results of a survey by the National Opinion Research Center, which measured the public's understanding of the nature of science showed that the number of science and math courses improved participants' performance on the survey (Impey, 2017). An ongoing assessment of undergraduate science literacy that includes content knowledge demonstrated that the number of science courses taken was "the strongest predictor of overall score" but accounted for only 3% of the variance indicating the influence of mediating variables such as attitude toward science (Impey, 2017, p. 173). Nuhfer et al. (2016) found that significant gains could be made within a single semester ($M = 2.95$, $p < .01$) but performance on the SLCI was not significantly different based on the total number of science courses completed until a student completed three or more science courses (Nuhfer et al., 2016). This potentially conflates academic interest in science with number of science courses completed as students completing three or more science courses are exceeding the minimum number needed for non-science majors. The number of science courses an undergraduate student completes had a positive relationship with content-dependent science literacy scores as measured by the SLCI (Nuhfer et al., 2016). The number of science courses taken increases the time on task for students practicing science literacy skills. The identification of the effect of number of science courses completed on students' science literacy could justify an increase in the number of science courses required to graduate high school from a low number of two to four.

The science requirements of NCLB, combined with the depth of the NGSS, necessitate a different instructional model. According to *A Nation at Risk,* "secondary school curricula have been homogenized, diluted, and diffused to the point that they no longer have a central purpose" (National Commission on Excellence in Education, 1983, p. 21). Several schools have increased the amount of instructional time by extending the school days, providing interventions, and increasing the number of courses required for students to graduate. The average number of science courses taken in high school has increased from 2.8 in 1990 to 3.5 in 2009 (National Science Foundation, 2006). The adoption of NCLB and Race to the Top led states to increase the number of science courses required to graduate. In Texas, between 2000 and 2015, the number of credits required to graduate rose from three to four, which required an additional year-long science course. Identifying the number of science courses completed as a variable impacting science literacy has policy implications promoting successful completion of more science courses for high school graduation.

Attitude Toward Science

Attitudes are moderately correlated with higher achievement but are difficult to change (Ajzen & Fishbein, 1980; Shrigley, 1990). In the United States, the correlation between attitude toward science and content knowledge differs by gender. Men's attitude toward science correlates with knowledge of science ($r = 0.366$, $p < .001$), while the correlation is lower for women (r = 0.284, $p < .001$; Hayes & Tariq, 2001).

Based on the Simpson-Troost Attitude Questionnaire (STAQ), student attitudes toward science and the classroom accounted for up to 18% of the variance in student achievement t (Talton & Simpson, 1987). A predictive regression model demonstrated that among male chemistry and physics students, there was a positive correlation between interest in science and math ($r = .31$, $p < .01$; $r = 0.28$, $p < .01$, respectively). For female students, there was no correlation (Larson et al., 2014). In a meta-analysis of studies, Weinburgh (1995) demonstrated a link between attitude toward science and science achievement that was particularly prevalent amongst female. In a review of existing data in Ogun State Nigeria, Olatoye (2009) examined the relationship between anxiety, motivation for an exam, and achievement on science tests. He found that motivation and anxiety could accounted for 14.6% of the variance on the science achievement test ($R^2 = 0.146$, $p < .05$).

Owen et al. (2008) used a standard pre-test/post-test design for measuring student content knowledge and attitude with an afterschool program to measure the effectiveness of the program. The STAQ-R was used as a measure of science motivation while a content specific assessment was developed for the research. The after-school program consisted of an inquiry curriculum to engage students ($N = 134$) 85% of whom were African American and Hispanic 88% of whom were eligible for FARP programs. A multiple regression analysis was performed with student gains on the subscales of the STAQ-R as the predictor variable used to measure content knowledge gains. Variance in the content knowledge gains were best predicted by the self-directed efforts and peer perception subscales of the STAQ-R. There is evidence to support *science is fun* subscale is a suppressor variable that is correlated with self-directed effort and may mask some of the variance due to self-direction (Newell et al., 2015). The predictive validity demonstrates that scores on the MSLQ are moderately correlated with final course grade performance (Pintrich et al., 1993). The TIMSS has demonstrated a "clear positive relationship was observed between stronger liking of mathematics and science and higher achievement" (Beaton et al., 1998, p. 70).

Motivation, attitude and engagement are domain specific. Two factors in the model, task management and anxiety, are "more general across mathematics, English, and science" indicating some domains may have significant overlap (Green et al., 2007). This supports the need to measure motivation and attitude using domain-specific measures. Simpson and Troost (1982) conducted a longitudinal investigation of how individual, home, and school variables impact students' commitment to science and achievement in science. The investigation involved self-reports, reviews of existing records, interviews with students and teachers, and measures of variance in enrollment and teacher evaluations. Commitment to science was an overarching term used to include "interest, attitudes, values, and other affective behaviors of student" (Simpson & Troost, 1982, p. 765). Researchers measuring the role of attitude in a behavior need to measure attitude for that specific construct such as science motivation.

A metanalytic study to assess the relationship between knowledge and attitudes using European datasets (N = 193) found a positive correlation between that general science knowledge and generalized attitudes toward science (r = 0.14), but there was no correlation between general science knowledge and specific attitudes toward genetically modified organisms (Allum, Besley, Gomez, & Brunton-Smith, 2008). Science literacy is generalized science knowledge indicating the inclusion of a generalized attitude measure would significantly impact science literacy. Science attitude assessments considered for this study were generalized science attitude surveys indicating a potential correlation with the generalized measure of science literacy.

Excluded Variables

Hellmuth (2014) used the TOSLS as the outcome measure of science literacy while testing the predictive value of Magical Ideation, Religiosity, Conservative-Liberalism, and Social Cynicism. The researcher used a multiple regression model to determine that scriptural literalism, religiosity, and magical ideation were significantly and negatively correlated with science literacy (Hellmuth, 2014). The sample size ($N = 43$) limited the ability of the study to assess demographic characteristics and science literacy. Magical ideation, religiosity and scriptural literalism are all excluded from inclusion in this study as they do not apply to public schools.

Chapter 3

METHOD OF PROCEDURE

The purpose of this chapter is to explain the research methodology for this quantitative study. The aim of this study was to assess factors that affect science literacy and use those factors to generate a predictive model. This approach helps quantify the factors that influence students' science literacy so that stakeholders can focus their efforts on improving it. The research plan includes the design of the study, description of the participant population, data collection, instrumentation, and data analysis procedures.

Design of Study

Science literacy is a complex construct that includes content knowledge, affective, and demographic factors. In this correlational study, using multiple regressions analyses, secondary data was reviewed and combined with information from two assessments to generate a predictive model of science literacy for high school students.

High School students from grades 10 to 12 completed the Test of Science Literacy Skills (TOSLS) as a measurement of science literacy during their existing class period. During the next class period, the Behavioral, Related Attitudes, and Intentions Towards Science (BRAINS) assessment was administered as a measure of attitude toward science. Students' academic records were used to gather data for the remaining variables. The model includes demographic, content knowledge, and affective measures reclassified as distal, opportunity, propensity, and demographic factors to generate a hierarchal multiple regression model.

A review of student records provided demographic information about the participants. Demographic variables previously correlated with science achievement or science literacy and routinely collected by schools including Gender, Ethnicity, English Language Proficiency, and Economic Need were included in the model.

A review of records provided the content knowledge variables for the model. Content knowledge variables previously correlated with science achievement or science literacy and routinely collected by schools include Number of Science Courses Completed, scores on State of Texas Assessment of Academic Readiness (STAAR) English Language Arts 1 (ELA 1), STAAR Algebra 1, and STAAR Biology. Student records provided students' Grade Point Average (GPA) on a 4.0 scale to minimize the impact of course selection on GPA. Grade Point Average was collected from educational records because of lack of accuracy of self-reported GPA scores (Kuncel, Credé, & Thomas, 2005). GPA was classified as an affective or distal variable because course selection, goal orientation (D'Agostino & Powers, 2009; Dweck, 1986), and student motivation (Crede & Phillips, 2011) affect GPA.

Sample Selection

The participant pool consisted of a convenience sample selected from a single suburban high school outside a major metropolitan area in Texas. Students enrolled in this school have historically performed well on state assessments; the school has received an A rating in its most recent evaluation (TAPR, 2018). Graduating students have not achieved the expected level of post-secondary success as evident by a decreasing percentage of students remaining enrolled in a college or university two years after graduation. Data provided by the ACT suggests that students are least prepared for college-level science with only 56% of students ready for college-level biology, compared to 79% prepared for college-level English composition (ACT, 2015). The percentage of students that attend a post-secondary institution is 64.8%; however, four-year longitudinal data from the National Student Clearinghouse suggests that only 47.5% of students who attend post-secondary institutions graduate (National Student Clearinghouse, 2015).

The demographics of the student population for this high school differ from TX state and national averages. Compared to the Texas state average, this suburban school has a significantly lower number of students identified with Economic Need, Limited English Proficiency, and ethnic minorities as seen in Table 3.

Table 3

Demographic Comparison Between the Students at the Research Site and Texas State Average

Group	Texas Population	Site Population
Economic need	47.7%	14.2%
Hispanic	49.2%	16.2%
African American	12.6%	5.3%
Limited English proficiency	5.3%	1.8%

Note. Date derived from Texas Academic Performance Report. (2018, December). Retrieved from https://rptsvr1.tea.texas.gov/perfreport/tapr/2018/srch.html?srch=C

The participant pool was limited to students for whom all information for the model was available, including Gender, Ethnicity, Economic Need, Limited English Proficiency, GPA, STAAR ELA 1, STAAR Algebra 1, STAAR Biology, Number of Science Courses Completed, and BRAINS score. This limited the participant sample to seniors, juniors, and a subsection of the sophomore class that had completed Biology before the start of the 2019-2020 school year. Freshmen were excluded from the model for a lack of information on their STAAR scores. A total of 1279 students were eligible to participate in the study.

Data Collection

Preparation for this study began in the summer of 2016, when a pilot study was conducted to assess the feasibility of this project. The TOSLS, a modified version of the STAAR Chemistry assessment, and the Motivated Strategies Learning Questionnaire (MSLQ), were administered to students enrolled in an advanced chemistry course. Performance on the TOSLS by these students (*M* = 63%) fell within the established range of scores (*M* = 44%–85%) established for post-secondary institutions (Chandler, 2017; Gormally et al., 2012). The MSLQ provided general information about motivation and learning style but was not designed to measure attitude toward science. The lack of information about science limited the usefulness of the MSLQ, so it was substituted with the BRAINS assessment. The BRAINS assessment explicitly measures attitude toward science and the relationship between attitude and observable behavior instead of generalized motivation (Summers & Abd-El-Khalick, 2018).

Student achievement (*M* = 76%) on a selection of questions from the released STAAR Chemistry exam was high but lacked comparable data due to the cancellation of the STAAR Chemistry assessment before widespread implementation (Chandler, 2017; Texas House Bill 5, 2013). Nationwide, 30% of students never take Chemistry, which limits the applicability of chemistry content knowledge to the general population (OECD, 2007; Roberts, 2007). In this study, the Chemistry STAAR scores were replaced with STAAR Biology scores as science content knowledge due to an expanded participant pool and the availability of a standardized instrument for Biology content knowledge. Performance on Algebra 1 and ELA 1 were included to reflect the multidimensional nature of science literacy (Holbrook & Rannikmae, 2007). English Language Arts (Tong et al., 2014) and Algebra (Wang & Ma, 2016) have also been linked to science achievement.

Prior to implementation, approval to conduct this study was obtained from the Institutional Review Board at Texas A&M University-Commerce. After obtaining site permission, a meeting with teachers was conducted to explain their role in the study and to obtain consent to use their classroom. Teacher permission forms were collected in the Fall 2019 semester. Teachers scheduled two, 35-minute sessions to administer the BRAINS survey and the TOSLS assessment. The modified block schedule resulted in the administration of assessment with 24 or 48 hours of time passing between each assessment. The limitations on the participant pool minimized the potential loss of instructional time on subjects with high-stakes accountability assessments.

During the first week of the spring semester, the researcher explained the scope of the study to students and provided both parental consent and student assent forms. Students were provided an incentive to return their consent and assent forms through entry into a raffle. Five winners were selected from those returning forms who won $20 gift card to either Apple Itunes or Google Play stores. Students were provided a two-week window to return consent and assent forms prior to data collection.

Teachers administered the assessments between January 20th and January 31st. Teachers and students were given the URL to the Google Forms versions of the TOSLS and BRAINS assessments and access to Google Chromebooks. The BRAINS survey was completed first class period with TOSLS administration the following class period. The TOSLS was administered after the BRAINS survey to minimize the impact a difficult assessment (*M* = 62%) may have on a students' attitude toward science (Chandler, 2017). Both assessments were completed in class with the support of the teacher to aid with technical difficulties. Data was collected directly from the Google Form into a Google Sheet, scored using established formulas including reverse scoring, and exported to Microsoft Excel with educational records review information.

Instrumentation

Attitude Toward Science was measured directly using a Google Form version of the BRAINS assessment. The outcome variable, science literacy, was directly assessed using a Google Form variation of the TOSLS.

Attitude Toward Science

The BRAINS assessment was selected as the most appropriate instrument to measure attitude toward science because of its predictability of the theoretical framework, modern psychometric validity, updated readability, and target participant age. The instrument consists of 30, 5-point Likert-scale items that measure attitude toward science. There are five subscales within the instrument including *attitude toward science, intention to pursue science, beliefs about behavior, control beliefs, and normative beliefs* (Summers & Abd-El-Khalick, 2018). Attitude is a contextual construct requiring the assessment target the specific area of interest (Green et al., 2007). The BRAINS assessment, unlike the MSLQ, specifically measures attitude toward science. The development of the BRAINS assessment was based onthe best items selected from seventeen existing science attitude assessments including the Attitude Toward Science in School Assessment (ATSSA) and Simpson-Troost Attitude Questionnaire (STAQ-R).

The BRAINS survey is based on the theory of reasoned action and planned behavior (TRAPB), a model that predicts behaviors (Ajzen, 2012; Sheppard et al., 1988). The TRAPB model is an extension of expectancy-value theory and accounts for the impact of cognitive mediators in predicting behavior. The BRAINS assessment fits the model outlined by TRAPB with a root mean square error of approximate values ($RMSEA < .05$), which represents a good fit

(Kelly & Lai, 2011). According to the model, BRAINS is a good fit for TRAPB, (*RMSEA* = .04). A comparative fit index (CFI) test resulted in a value of .95, confirming a good fit to the model. The BRAINS assessment fits the theoretical framework for the TRAPB model and is likely to accurately predict students' intentions and behavior.

The BRAINS assessment has established psychometric validity without the assumption of normality. Confirmatory factor analysis has demonstrated the instrument loads on to a unitary construct (Summers & Abd-El-Khalick, 2018). Internal reliability was assessed by subscale with estimates ranging from $\alpha = .70$ to $\alpha = .91$ (Summers & Abd-El-Khalick, 2018). The BRAINS assessment has face validity with experts, which is a good fit to a robust model. It also demonstrates strong reliability and validity. The target population of the BRAINS are students of 5th to 10th grades, but it can be used with students of higher grades (Summers & Abd-El-Khalick, 2018). The readability of the BRAINS, its modern psychometric approach, and its updated standards of clarity also make it the most appropriate instrument for this study. The BRAINS assessment lacks widespread usage in the literature because it was released in 2018.

To ease distribution and data acquisition from multiple classrooms over multiple days, with the permission of the instrument developer, the instrument was converted from a paper assessment to a Google Form. The conversion of paper assessments has shown negligible impact on the responses when converted to electronic surveys (Coke, 2014). A copy of the BRAINS assessment is included in Appendix A. Scores for each item were reported from 0–4 with a total score ranging from 0–120. Subscale scores were collected despite the fact that direct comparison between categories was not possible because of the variation in the total number of questions in each category.

Science Literacy

The TOSLS was the most appropriate instrument to assess science literacy. The TOSLS is a 28-question multiple choice assessment that measures science literacy using nine skills divided into two domains. The two domains of the assessment include methods of inquiry and quantitative analysis. The TOSLS was originally developed for use at undergraduate institutions but has similar mean performance when used in a high school setting (Chandler, 2017). All items include biology as the context for the question, but do not require content knowledge to answer. The independence of content knowledge from science literacy aligns with the operational definition used in this study. Mean scores range from M = 43% to M = 85% (Gormally et. al, 2012) with advanced high school students scoring 62% (Chandler, 2017).

The TOSLS has content validity and reliability. Content validity for the TOSLS was achieved by consulting a panel of experts. The TOSLS also aligns with state (Texas Essential Knowledge and Skills for Science, 2018) and national (NGSS Lead States, 2013) standards for science literacy. The TOSLS has acceptable reliability ($\alpha = 0.748$) with exploratory factor analysis demonstrating a single factor of science literacy as the likely cause of variability in scores. The TOSLS has been used to measure progress within semester-long university courses (Waldo, 2015) and differences in student performance between groups taught different curricula (Auerbach & Schussler, 2017).

A link to access the electronic version of the TOSLS was provided to the teachers (See Appendix B for a reproduction of the instrument) that provided students logged into their district account access to the form. Scores were collected as the total number correct responses ranging from 0– 28 with each skill subscale score also tabulated. The number of items correct in each subscale vary from 1 question in subscale 5 to 5 questions in subscale 2.

Records Review

Students' demographic and knowledge information were obtained from existing educational records to minimize both incidental and self-report errors. Cook and Campbell (1979) demonstrated that participants are influenced by their perception of the researcher's expectations as well as by what would reflect positively on the participant. Data regarding the demographic predictor variables of Gender, Ethnicity, English Language Proficiency, and Economic Need were obtained from the school registrar using the Skyward Student Information System. Demographic information is self-reported by the parents at enrollment and is verified annually. Categorical variables were coded as shown in Table 4.

Table 4

Categorical Variable Coding for Demographic Variables

Variable	0	1	2	3	4	5
Gender	Female	Male				
English proficiency	Limited	Proficient				
Economic status	Not in economic need	Economic Need				
Ethnicity	American Indian	Asian	Black	Hispanic	Two or More	White

Transcripts were reviewed for the Number of Science Courses Completed for high school credit. Credit was tabulated in half-course units with a semester average greater than 70 being counted as .5 on the student transcript. Anatomy and Physiology, Forensic Science, Medical Terminology, Engineering classes, and Health Science classes can be completed for either science or career and technology credit. Based on post-hoc analysis, these courses were separated and classified as an additional category for potential inclusion in Number of Science

Courses Completed. The Number of Science Courses Completed was estimated to range of 0 to 10 with most students earning 0.5 to 4 credits.

STAAR assessment results were obtained in a review of educational records. STAAR tests are administered annually from 3rd grade through 11th grade. High school students are required to successfully complete five STAAR exams: English 1, English 2, Algebra 1, Biology, and U.S. History. Participants' STAAR scores were obtained in Biology, ELA 1, and Algebra 1 as an existing standardized measure of content knowledge in these domains. Data was obtained using the AWARE database within Eduphoria. STAAR scores are available in raw number correct, percentage correct, and as a scaled score. Variation in the number of questions and range of scaled scores necessitated percent correct information be collected. This allowed for the direct comparison of the three content knowledge variables in the model. This operational definition limited the range of potential scores to between 0–100%. The range of continuous variables collected is shown in Table 5.

Table 5

Range of Continuous Variables From Educational Records Review

Variable	Minimum	Maximum
Science courses completed	0	10
STAAR English 1	0%	100%
STAAR Biology	0%	100%
STAAR Algebra 1	0%	100%
GPA	0	4.0

GPA was collected from transcripts using the traditional 4.0 scale. The 4.0 scale was selected to avoid conflation between course selection and academic performance. GPA was limited to a range of 0 to 4.0

Treatment of Data

The researcher coded all information into a single Excel document that contained personally identifiable information including names and student identification numbers. Names and student identification numbers were used to match data and as a fidelity check prior to analysis. The researcher conducted a fidelity check by randomly selecting 10% of the participants (n = 45) and verifying that all variables were entered correctly. After the completion of this study, all data was transferred to Dr. Naizer. Identifiable information was removed from the spreadsheet before porting to SPSS for analysis. IBM's SPSS Version 26 was used for all subsequent analysis. Significance testing was conducted at .05, which was consistent with literature on science literacy (Germann, 1988; Gormally et al., 2012; Nuhfer et al., 2016; OECD, 2016).

TOSLS data was assessed for outliers by generating a histogram of scores and calculating interquartile ranges to reinforce the assumption that participants actively tried their best (Nuhfer et al., 2016). TOSLS scores were plotted in a histogram with scores falling outside the normalized curve identified as possible outliers. Interquartile ranges were calculated by determining the difference in weighted average for the first and third quartiles, multiplying the value by one and one half. This value was subtracted from the average of the first quartile to determine the lower bound of acceptable data points and added to the average of the third quartile to determine the upper bound. Data points classified as outliers and information that with missing fields was removed (J. W. Osborne & Overbay, 2004).

The dataset was analyzed for normalized distribution and homogeneity of variance. A Kolmogorov-Smirnov test of normality was conducted to assess the normality of distribution for TOSLS scores. The K-S test demonstrates normality when $p < .05$ as it tests the hypothesis that the groups are not normalized (Field, 2013). The assumption of normality is used to ensure scores represent a larger population. The homogeneity of variance for categorical variables was assessed using Levene's test of variance. This measures the difference between the mean and individual scores for categorical variables. Equal variance between categories is assumed when $p > .5$, allowing the samples to be compared using a *t*-test or ANOVA (Field, 2013).

A preliminary analysis of variables was conducted to assess the significance of each predictor variable, which allowed a direct comparison with existing literature. Categorical variables were assessed using a *t*-test or ANOVA while continuous variables were analyzed using Pearson correlations. An independent samples *t*-test was conducted using the descriptive predictor variables: Gender, English Language Proficiency, and Economic Need. The number of categories in Ethnicity necessitated an ANOVA analysis. A Pearson correlation was conducted to determine statistical significance and correlation between predictor variables and TOSLS scores. Correlation and significance between predictor variables was examined for strength and deviation from the established literature. The statistical significance was used to verify findings of previous research but did not exclude variables in the hierarchal multiple regression model.

Covariance of variables were assessed prior to inclusion in the model. A preliminary correlation matrix was generated to support the lack of collinearity between variables. Variables with a correlation coefficient, $r > .9$ warrant further exploration for autocorrelation. A large correlational value may indicate two variables warrant consideration as a single variable model. A Durbin-Watson test was conducted to verify a lack of autocorrelation between factors (Field, 2013). After the model was generated, collinearity variable inflation factor analysis was conducted to examine regression coefficients for collinearity.

Variables were reclassified from content knowledge, affective, and demographic factors into distal, opportunity, propensity and demographic factors. This was done to examine the impact of each variable in a chronological manner. Distal variables are factors that "operate earlier in time and explain the emergence of opportunities and propensities" (Byrnes & Miller, 2007, p. 602). Distal variables, such as English Proficiency, can disproportionately impact science literacy due to the duration of its effect. This required that distal factors be included first in the hierarchal multiple regression model because they represent up to 43% of the variance in achievement, followed by opportunity factors, propensity factors, and other variables of interest such as demographic factors (Byrnes & Miller, 2007). Distal variables in this study include Economic Need, English Language Proficiency, and GPA. Opportunity variables are factors that influence access to the outcome variable with Number of Science Courses Completed being the only opportunity variable included in this study. Propensity variables measure the tendency of a participant to be successful and included Attitude Toward Science, STAAR Biology, STAAR Algebra 1, and STAAR ELA 1. Demographic variables include Gender and Ethnicity. Previous research demonstrated the nearly 14% of variance in achievement was reduced to 3% once distal, opportunity, and propensity variables had been accounted for (Byrnes & Miller, 2007). Variable reclassification is shown in Table 1.

A multiple regression model was generated using hierarchal entry. Variables were entered into multiple regression analysis in the following order: Economic Need, English Language Proficiency, GPA, Number of Science Courses Completed, BRAINS Score, STAAR Biology, STAAR Math, STAAR ELA, Gender, and Ethnicity shown in equations below.

$$\text{Science Literacy} = \beta_1 EN + \beta_2 EP + \beta_3 GP + \beta_4 SC + \beta_5 AS + \beta_6 SB + \beta_7 SA + \beta_8 SE + \beta_9 GE + \beta_{10} ET \quad (1)$$

Distal Factors

EN – Economic Need

EP – English Proficiency

GP – Grade Point Average

Opportunity Factors-

SC – Science courses completed

Propensity Factors-

AS – Attitude Toward Science

SA – STAAR Algebra 1

SB – STAAR Biology

SE – STAAR English Language Arts 1

Demographic Factors -

GE – Gender

ET – Ethnicity

$$\text{Science Literacy} = \beta_A DI + \beta_B OP + \beta_C PR + \beta_D DE \quad (2)$$

DI - Distal Factors

O - Opportunity Factors

P - Propensity Factors

DE - Demographic Factors

The hierarchal, multiple regression model was developed with science literacy as the outcome variable. All variables were assessed with significance value of $p < .05$.

Unstandardized β values for each parameter in the models were determined with the potential impact of each variable on science literacy.

Summary

The goal of this chapter was to outline the research method used to answer the research questions. A discussion of the procedure, participant population, instrumentation, and treatment of the data was outlined. The extension of expectancy-value theory to the TRAPB model for predicting behavior was used to develop a model for predicting science literacy, as measured by the TOSLS test, using Economic Need, English Language Proficiency, GPA, Number of Science Courses Completed, BRAINS Score, STAAR Biology, STAAR Math, STAAR ELA, Gender, and Ethnicity. The goal of Chapter 4 is to provide the study results and to demonstrate that the methodology described in Chapter 3 was implemented.

Chapter 4

PRESENTATION OF DATA

The purpose of this study was to develop a hierarchal multiple regression model of factors commonly collected by schools to predict high school students' science literacy. These factors were divided into distal factors that "operate earlier in time and explain the emergence of opportunities and propensities" (Byrnes & Miller, 2007, p. 602), opportunity factors that influence students' ability to access the information assessed, propensity factors as measures of natural tendency, and demographic factors. Individual factors were assessed for significance and impact on science literacy prior to inclusion in the multiple variable regression model. The multiple regression model was generated using data from all collected variables.

Sample

Teachers (n = 12) gave permission for the study to be conducted in their classroom resulting in 1279 students eligible to participate in the study. A total of 445 students returned all required permissions to participate in the study with 3 students opting out. Student names and identification numbers were matched with student assent forms and parental consent forms, TOSLS scores, BRAINS scores, and review of educational records from the Eduphoria and Skyward Information Systems. Data from 66 students was removed because of missing information required to generate the multiple regression model, leaving a total of 376 participants in the study (see Table 6).

A fidelity check was conducted to verify the accuracy of data from primary sources. A random number generator was used to select 10% of the participants and verify the accuracy of their information in the Model Development set. This resulted in verification of information for 45 participants with three errors found resulting in an effective error rate of less than 1%. Errors

Table 6

Reasons for Data Removal Prior to Analysis

Removal Reason	*n*	Percentage
Students who opted out	3	0.7%
Missing BRAINS	3	0.7%
Missing STAAR test	7	1.6%
Missing TOSLS	12	2.7%
Missing multiple	43	9.7%
Total	69	15.5%

included one transcription error, an incorrect TOSLS score, and incorrect assessment of a career and technology education credit. The error rate was found to be less than that found in a study comparing accuracy and cost of data collection through manual transcription of data with computerized collection (Weber, Yarandi, Rowe, & Weber, 2005). Participant name and identification numbers were removed and transferred to SPSS Version 26 for further analysis.

A histogram of TOSLS scores with normalized distribution was generated to aid the detection of outliers (see Figure 1; Field, 2013). Data outliers occurred when violations of the assumption of best and truthful answers occurred as outlined in limitations of Chapter 1. Interquartile ranges were calculated, and all scores fell within the interquartile range. No data points were removed from the data set for testing, which resulted in a total of 376 participants being included in Model Development data set.

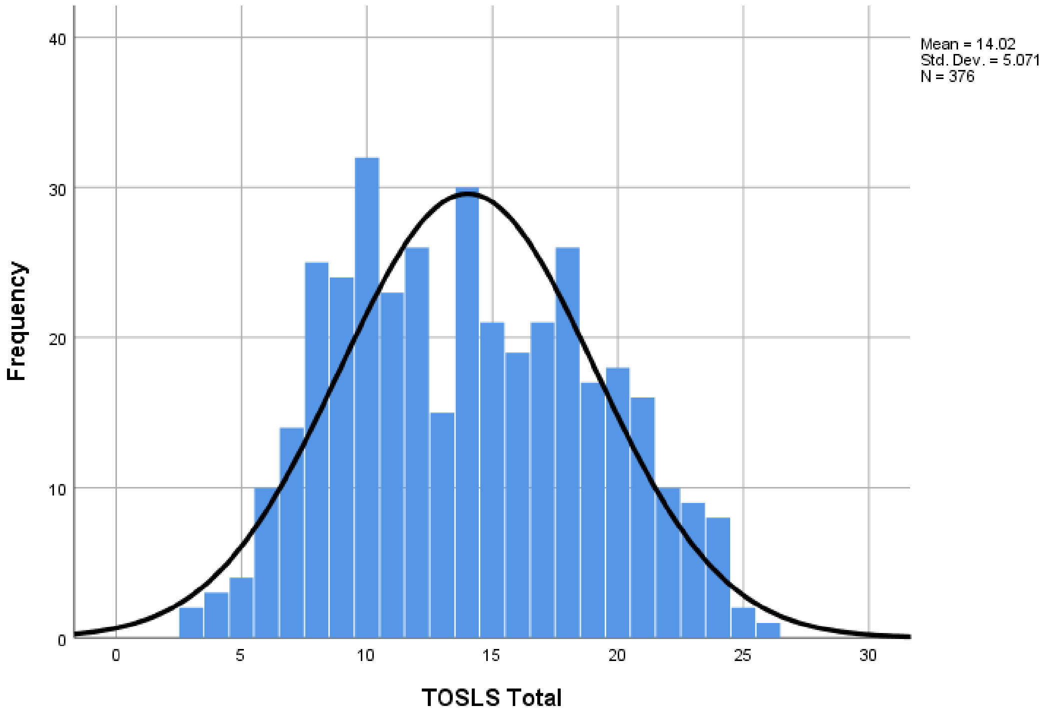


Figure 1. Histogram of TOSLS scores to identify outliers.

A one sample Kolmogorov-Smith (K-S) test of normality was conducted to verify the assumption of normality in TOSLS scores for the participant population. A K-S test was used to determine if the participant population is representative of a larger population (Field, 2013). The null hypothesis that the participant sample is different from the larger population was assessed at the threshold of $p < .05$. Significance at $p = .000$ supports the assumption of normality that the participant population represents a larger population.

Levene's test was used to assess homogeneity of variance for the categorical variables of Economic Need, English Proficiency, Gender, and Ethnicity. Levene's test measures the difference between the mean and individual scores for categorical variables (Field, 2013). Equal variance for TOSLS can be assumed when $p > .05$. The TOSLS scores were not significantly different in variation for gender $F(1, 374) = 0.260$, $p = .610$, ethnicity $F(5, 370) = 1.036$, $p =$

.396, economic need $F(1, 374) = 0.353$, $p = .553$, and English proficiency $F(1, 374) = 0.036$, $p = .849$. This indicates homogeneity of variance and allows for the comparison of means using a *t*-test or ANOVA analysis.

Science Literacy Performance

Performance on the TOSLS fell within the range of variation by undergraduate institution established by Gormally in 2012 from $M = 42.5\%$ to $M = 84.95\%$. The participant population was at the lower end of this range, ($M = 50.06\%$, $SD = 17.86\%$) with raw scores shown in Table 7, which is consistent with the use of an undergraduate assessment in a high school setting. The mean performance, similar standard deviation, and normalized distribution justify the TOSLS as an appropriate instrument for secondary high school students.

Preliminary Analysis

A preliminary analysis of each predictive factor was conducted to assess the relationship between each predictive factor and science literacy for comparison with established literature. This step was deemed necessary because in this study, the factor of science achievement was used as a proxy for science literacy when the operational definition for science literacy varied or information was not available. Descriptive statistics, mean performance, standard deviations, and normality of variables were included in the preliminary analysis by predictive factor. Each predictive factor was assessed for statistical significance as an independent factor impacting science literacy. The categorical variables Economic Need, English Proficiency, and Gender were assessed using an independent samples *t*-test. Ethnicity was assessed using an ANOVA test. The continuous variables STAAR Algebra 1, STAAR Biology, STAAR ELA 1, GPA, Number of Science Courses Completed, and BRAINS assessment were correlated and assessed for significance using the Pearson *r* statistic.

Table 7

Mean Performance on Test of Science Literacy Skills Assessment by Skill

Variable	*n*	*Minimum*	*Maximum*	*M*	*SD*
TOSLS Total	376	3	26	14.02	5.071
Skill #1	376	0	3	1.93	0.963
Skill #2	376	0	5	2.13	1.243
Skill #3	376	0	3	1.65	0.893
Skill #4	376	0	4	1.91	1.048
Skill #5	376	0	1	0.37	0.484
Skill #6	376	0	4	2.27	1.226
Skill #7	376	0	3	1.59	0.986
Skill #8	376	0	3	1.15	0.836
Skill #9	376	0	2	1.02	0.771

Distal Variables

Distal variables have the potential to impact educational outcomes over a greater period of time while creating longitudinal knowledge deficits. The classification of Economic Need and English Proficiency as distal variables necessitated that students be classified as participants in a program if they ever participated in that program. The distal variables were included first in the model based on the potential long-term impact on outcomes (Byrnes & Miller, 2007). Economic Need was the first variable to be included in the distal variable model with primacy in impact in the literature, followed by English Proficiency, and GPA.

Economic Need. Economic Need was defined as participation in the free and reduced lunch program at any time while enrolled in public school. The income necessary to qualify varies annually but is defined as up to 1.8 times the federal poverty guidelines. Eligibility was accessed from the Skyward Information System from the category of Food Services. Students identified with Economic Need (n = 77) represented 20.5% of the population. There was a significant difference in mean TOSLS scores for students classified as in economic need (M = 12.13, SD = 4.900) and those without (M = 14.50, SD = 5.008, $t(374) = 3.722$, $p < .000$ (see Table 8).

Table 8

Independent Samples t-Test Between TOSLS Score and Students' Economic Status

Economic status	*n*	*M*	*SD*	*t-cal*	*df*	*p*
Need	6.9%	12.13	4.900	3.772	374	0.000
No need	20.8%	14.50	5.008			

English Proficiency. English Proficiency was defined as participation in the Texas English Language Proficiency Assessment System (TELPAS) without differentiation between proficiency levels. This information was accessed from the Skyward Information System under LEP/Bilingual. Only 6.9% of the participants (n = 26) were classified with Limited English Proficiency. This included students that have been dismissed from the LEP program by scoring Advanced High on previous TELPAS assessments. There was not a significant difference in mean TOSLS scores for students of Limited English Proficiency (M = 14.38, SD = 4.883) and proficient (M = 13.99, SD = 5.091); $t(374) = -.384$, $p = .701$ (see Table 9).

Table 9

Independent Samples t-Test Between TOSLS Score and Students' English Proficiency

English Proficiency	*n*	*Mean*	*SD*	*t-cal*	*df*	*p*
Limited	26	14.38	4.883	-0.384	374	0.701
Proficient	350	14.38	4.883			

Grade Point Average. Grade Point Average (GPA) is a continuous variable ranging from 0 to 4.0 and represents the average of academic achievement for all coursework completed by the end of Fall 2020. GPA reflects effort and the ability to do well in school and is not considered a measure of intelligence. GPA scores were obtained through a review of education records in the Skyward Information System. The GPA for participants in the sample was (M = 3.389, SD = .504) as shown in Table 10. As shown in Table 11, GPA was positively correlated with TOSLS performance, $r(374) = .467$, $p = .000$.

Distal Model. A multiple regression model was generated using distal factors and science literacy to quantify the impact of these factors. Distal factors were moderately correlated $r(374) = .475$, $p = .000$ while accounting for 22% of the variance in science literacy as measured by the TOSLS.

Table 10

Mean and Range of Grade Point Averages

Variable	*n*	Minimum	Maximum	*Mean*	*SD*
GPA	376	1.704	4.000	3.39	0.503

Table 11

Correlation of TOSLS Total With Grade Point Average

Measure		TOSLS Total	Grade Point Average
TOSLS Total	Pearson Correlation	1	.467**
	Sig. (2-tailed)		.000
	N	376	376

Note.**Correlation is significant at the 0.01 level (2-tailed).

Opportunity Variable

Number of Science Courses Completed is the only opportunity variable in this study. The Number of Science Courses Completed was obtained from student transcripts through a review of education records in the Skyward Information System. The average number of science courses completed by participants was (M = 2.416, SD = .799) as shown in Table 12 along with range. The Number of Science Courses Completed was positively correlated with TOSLS performance, $r(374) = .202$, p = .000 as shown in Table 13. A multiple regression model was not conducted for opportunity variables as Number of Science courses completed was the only variable included. This accounted for 3.8% of the variance in science literacy.

Table 12

Mean and Range for Number of Science Courses Completed

Variable	*n*	Minimum	Maximum	*M*	*SD*
Science Courses	376	0.5	4.5	2.416	.799

Table 13

Correlation of TOSLS Total With Number of Science Courses Completed

Measure		TOSLS Total	Grade Point Average
TOSLS Total	Pearson Correlation	1	.467**
	Sig. (2-tailed)		.000
	N	376	376

Note. ** Correlation is significant at the 0.01 level (2-tailed).

Propensity Variables

Propensity variables measure the tendency of an individual to be successful on the science literacy assessment. Propensity variables included in this study are attitude toward science, STAAR Biology, STAAR Algebra 1, and STAAR ELA 1. Propensity variables are included after distal and propensity factors in this model.

Biology content knowledge. State of Texas Assessment of Academic Readiness (STAAR) Biology scores were obtained as a percentage of correct questions through a review of educational records in the Eduphoria system. STAAR Biology represents an assessment of student science content knowledge which includes science literacy. The average STAAR Biology score for student participants in the sample was (M = 81.90, SD = 13.298) with a range from 38% to 100% as shown in Table 14. As illustrated in Table 15, students' STAAR Biology score was positively correlated with TOSLS performance, $r(374) = .146$, $p = .005$.

Table 14

Mean and Range Summary for STAAR Biology Scores

Variable	*n*	Minimum	Maximum	*M*	*SD*
STAAR Biology	376	38	100	81.90	13.298

Table 15

Correlation of TOSLS Total With STAAR Biology Scores

Measure		TOSLS Total	STAAR Biology
TOSLS Total	Pearson Correlation	1	.146**
	Sig. (2-tailed)		.005
	N	376	376

Note. ** Correlation is significant at the 0.01 level (2-tailed).

Algebra content knowledge. STAAR Algebra 1 scores were obtained as a percentage of correct questions through a review of educational records in the Eduphoria system. STAAR Algebra 1 scores represent an assessment of student mathematical content knowledge with overlap to quantitative literacy. The average STAAR Algebra 1 score for student participants in the sample was (M = 80.48, SD = 15.24) as shown in Table 16. The STAAR Algebra 1 score was positively correlated with TOSLS performance, $r(374) = .434$, $p = .000$ as shown in Table 17.

Table 16

Mean and Range for STAAR Algebra 1 Scores

Variable	*n*	Minimum	Maximum	*M*	*SD*
STAAR Algebra	376	9	100	80.48	15.24

Table 17

Correlation of TOSLS Total With STAAR Algebra 1 Scores

Measure		TOSLS Total	STAAR Algebra
TOSLS Total	Pearson Correlation	1	.434**
	Sig. (2-tailed)		.000
	N	376	376

Note. ** Correlation is significant at the 0.01 level (2-tailed).

English language arts content knowledge. STAAR English Language Arts (ELA) 1 scores were obtained as a percentage of correct questions through a review of educational records in the Eduphoria system. STAAR ELA 1 scores represent an assessment of English Language content knowledge with overlap to general literacy. The average STAAR ELA 1 score for student participants was shown in Table 18 to be (M = 80.85, SD = 10.698) for the participant population. The STAAR ELA 1 score was positively correlated with TOSLS performance, $r(376) = .498$, $p = .000$ as shown in Table 19.

Table 18

Mean and Range for STAAR English Language Arts 1 Scores

Variable	*n*	Minimum	Maximum	Mean	SD
STAAR ELA 1	376	25	100	80.85	10.698

Table 19

Correlation of TOSLS Total With STAAR English Language Arts 1

Dependent Var	Measure	TOSLS Total	STAAR English
TOSLS Total	Pearson Correlation	1	.498**
	Sig. (2-tailed)		.000
	N	376	376

Note. ** Correlation is significant at the 0.01 level (2-tailed).

Attitude toward science. Attitude toward science was measured using the Behavioral, Related Attitudes, and Intentions Towards Science (BRAINS) assessment. The BRAINS assessment was administered to students electronically. The BRAINS assessment consists of a single instrument with five subscales. Table 20 illustrates students' overall performance on the BRAINS as well as performance by subscale. The average BRAINS score for student participants in the sample was (M = 72.69, SD =19.07). The BRAINS score was positively correlated with TOSLS performance, $r(376) = .385, p = .000$ as shown in Table 21.

Propensity model. Propensity factors, which are STAAR Biology, STAAR Algebra, STAAR English Language Arts, and BRAINS assessment, were included in a multiple regression model with science literacy. Propensity factors were positively and moderately correlated with science literacy $r(371) = .601$, $p = .000$ while accounting for 35.5% of the variance in TOSLS performance.

Table 20

Mean and Range for Behaviors Related Attitudes and Intention Toward Science by Subscale

Variable	*N*	Min	Max	Mean	SD
Attitudes toward science	376	0	24	14.61	5.713
Intentions to pursue/engage	376	0	24	11.85	6.390
Beliefs about behavior	376	10	33	24.97	4.272
Control beliefs	376	0	24	16.14	4.781
Normative beliefs	376	0	12	5.11	3.018
TOTAL	376	17	114	72.69	19.072

Table 21

Correlation of TOSLS Total With BRAINS Total

Dependent Var	Measure	BRAINS Total
TOSLS Total	Pearson Correlation	.385**
	Sig. (2-tailed)	.000
	N	376

Note. ** Correlation is significant at the 0.01 level (2-tailed).

Demographic Variables

Demographic variables were the last to be included in the stepwise regression model due to their limited impact after accounting for distal, opportunity, and propensity variables (Byrnes & Miller, 2007). Demographic variables in this study include gender and ethnicity.

Gender. Gender was treated as a dichotomous variable and obtained through the Eduphoria system. Two hundred and fifteen students (57.2%) were female. As illustrated in Table 22, there was a significant difference in the mean TOSLS scores for female (M = 13.53, SD = 5.02) and male (M = 14.67, SD = 5.08) participants, $t(374) = -2.178$, $p = .030$.

Table 22

Independent Samples t-Test Between TOSLS Score and Students' Gender

Gender	*n*	*M*	*SD*	*t-cal*	*df*	*p*
Female	215	13.53	5.02	-2.178	374	.030
Male	161	14.67	5.08			

Ethnicity. Ethnicity was obtained through a review of educational records in the Eduphoria system. Most students (n = 251) were classified as white, which represented 66.8% of the total participant population. The ethnicity of the participant population is illustrated in Table 23. As illustrated in Table 24, there was a significant difference in the TOSLS scores between ethnic groups $F(5,370) = 2.427$, $p = .035$.

Table 23

Frequency and Mean TOSLS Performance by Ethnicity

Ethnicity	*n*	*M*	*SD*
American Indian/Alaskan	3	16.33	5.508
Asian	16	17.31	4.771
Black/African American	20	13.80	5.268
Hispanic	66	12.94	4.503
Two or More Races	20	15.40	5.924
White	251	13.97	5.069

Table 24

ANOVA for Equality of Means Between TOSLS Score and Students' Ethnicity

Measure	Sum of Squares	*df*	Mean Square	*F*	Sig
Between groups	306.297	5	61.259	2.427	.035
Within groups	9337.607	370	25.237		
Total	9643.904	375			

Note. ** Correlation is significant at the 0.05 level (2-tailed).

The Bonferroni Correction measured the mean differences between groups post-hoc to determine significance between groups. The Bonferroni Correction adjusts the significance level by dividing the significance level by the number of comparisons (Field, 2013). Asian students (M = 17.31, SD = 4.771) significantly (p = .029) outperformed Hispanic students (M = 12.94, SD

= 4.503). The difference in means between other ethnic groups did not rise to the level of significance.

Ethnicity was converted from categorical variable with six options to six dummy binary variables in preparation for inclusion into a multiple regression model. This step was necessary as the numerical assignment of ethnicity does not represent a scale variable. Mean performance for each group was compared with the remainder of the population as shown in Table 25. The resulting change to the model equation is shown below with ethnicity entered as six dummy variables:

$$B_{10}ET = \beta_{11}AI + \beta_{12}AN + \beta_{13}BK + \beta_{14}HC + \beta_{15}TE + \beta_{16}WE \quad (3)$$

The expression shown above will reduce to a single beta coefficient and variable for every student. A student classified as Hispanic would be entered as:

$$B_{10}ET = \beta_{11}(0) + \beta_{12}(0) + \beta_{13}(0) + \beta_{14}HC + \beta_{15}(0) + \beta_{16}(0) \quad (4)$$

$$\beta_{9}ET = \beta_{14}HC \quad (5)$$

Table 25

Bonferroni Correction Mean Comparison of TOSLS Scores by Ethnicity

Ethnic Group	*n*	*M*	*SD*	*ΔM*	Sig
American Indian (AI)	3	16.33	5.508	2.336	.428
Asian (AN)	16	17.31	4.771	3.443	.008
Black (BK)	20	13.80	5.268	-.228	.845
Hispanic (HC)	66	12.94	4.503	-1.306	.057
Two or More Races (TE)	20	15.40	5.924	1.462	.210
White (WE)	190	14.27	5.232	.510	.330

Demographic model. A multiple regression model was generated for demographic factors and science literacy. Demographic factors were moderately correlated with science literacy performance $r(368) = .206$, $p = .0205$ while accounting for 2.4% of the variance.

Model Development

The research question sought to measure the impact of distal, opportunity, propensity and demographic factors to develop a model for science literacy. Multiple regression analysis was performed involving all variables for each classification to generate a distal model, opportunity model, propensity model, and demographic model. Models were examined for relative impact before being combined into a hierarchal regression model that accounted for all remaining significant predictive factors.

The overall question and subsequent parts include:

1. How much will variation in demographic, distal, propensity, and opportunity factors predict science literacy as measured by the TOSLS?
    a. What is the relationship between distal factors (Economic Need, English Proficiency, and Grade Point Average) and science literacy?
    b. What is the relationship between propensity factors (Attitude Toward Science and multi-subject content knowledge) and science literacy?
    c. What is the relationship between opportunity factors (# of science courses completed) and science literacy?
    d. What is the relationship between demographic factors (ethnicity and gender) and science literacy?

The categorical models were assessed for predictive impact by comparing the adjusted R^2 values to determine the predictive impact of each category of variable independent of other categories. A multiple regression model was developed using distal factors resulting in an adjusted value R^2 = 0.218 or 21.8% of the variance in TOSLS scores. Opportunity factors accounted for 6.5%, propensity factors 15.3% and demographic factors 0.9% of the variance of the variance in TOSLS scores.

A hierarchal, multiple regression model was developed using predictor variables with dummy variables substituted for ethnicity. Variables were added in the order Economic Need, Grade Point Average, Science Courses Completed, BRAINS Total, STAAR Biology, STAAR Algebra, STAAR English, Ethnicity, and Gender consistent with distal, opportunity, propensity model. Beta coefficient values for predictor variables are reported in Table 26. The model excluded economic need, English proficiency and STAAR Algebra 1 scores for not contributing to the predictability of TOSLS scores.

A correlation matrix was generated for all variables to assess collinearity of variables included. The correlation of predictive variables exceeding 0.9 may indicate that two variables are measuring the same construct (Field, 2013). The largest correlation coefficient between the variables was between STAAR English scores and STAAR Biology scores $r(374) = 0.708$, $p = .000$. The correlation was lower than the threshold supporting that each variable independently contributes to TOSLS scores. The correlation matrix also demonstrated that ethnicity was not significantly correlated with TOSLS scores, GPA, Science Courses Completed, BRAINS total score, STAAR Biology Score, and STAAR English score. Gender was not significantly correlated with Economic Need, STAAR Biology score, and STAAR Algebra 1 score. The Durbin-Watson test was calculated (2.050) to assess the autocorrelation of variables. A value of 2 is considered ideal while values with a range of 1–3 are considered acceptable providing support to the independent contribution of each variable. Collinearity was also assessed using

Table 26

Hierarchal Multiple Regression Coefficients for TOSLS Score

Variable	Unstandardized Coefficients		Stand Coeff	t	Sig.	95.0% Confidence		Collinearity Statistics	
	B	Std. E	Beta			Lower	Upper	Tol	VIF
b	-14.212	1.89		-7.519	0	-17.93	-10.495		
EN	-1.049	0.523	-0.084	-2.007	0.046	-2.077	-0.021	0.853	1.173
EP	0.936	0.852	0.047	1.098	0.273	-0.74	2.611	0.813	1.231
GP	0.668	0.639	0.066	1.045	0.297	-0.589	1.924	0.368	2.719
SC	1.772	0.254	0.279	6.971	0	1.272	2.272	0.924	1.083
SB	0.087	0.025	0.228	3.549	0	0.039	0.135	0.358	2.794
SE	0.119	0.029	0.251	4.126	0	0.062	0.176	0.4	2.497
SA	0.02	0.019	0.061	1.06	0.29	-0.017	0.058	0.448	2.233
AS	0.041	0.011	0.154	3.595	0	0.019	0.063	0.806	1.241
GE	0.807	0.421	0.079	1.918	0.056	-0.02	1.635	0.875	1.143
AI	2.511	2.202	0.044	1.141	0.255	-1.818	6.841	0.99	1.011
AN	1.231	1.034	0.049	1.191	0.235	-0.802	3.264	0.872	1.147
BK	0.858	0.913	0.038	0.939	0.348	-0.939	2.654	0.903	1.107
HC	-0.893	0.561	-0.067	-1.591	0.113	-1.997	0.211	0.833	1.201
TE	1.065	0.887	0.047	1.201	0.23	-0.678	2.809	0.959	1.043
WE	0.234	0.401	0.023	0.583	0.56	-0.555	1.023	0.944	1.059

the variance inflation factor (VIF) to quantify multicollinearity. The average VIF for the model was calculated to be 1.666, which was less than the 2.5, which requires further analysis for multicollinearity (Field, 2013).

The hierarchal model was assessed for significant changes to the predictability of the model. A P-P plot (see Figure 2) was generated to examine the skew of the predictive probability of the model with the observed probability demonstrating a normalized distribution. An examination of the predictability of the model was conducted. Case wise residuals indicate that 3.19% of TOSLS scores fell outside two standard deviations from the mean with a single case, 0.27%, outside of three standard deviations.

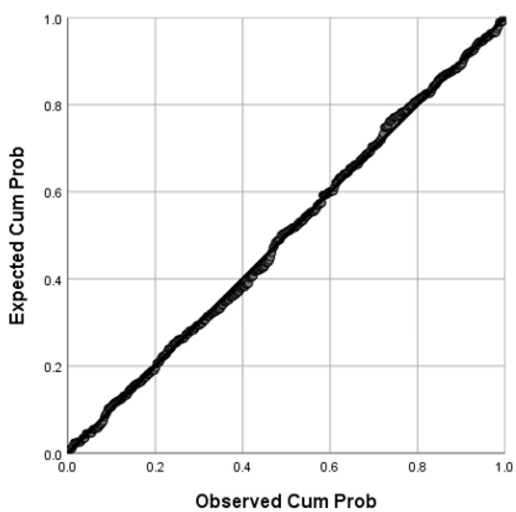


Figure 2. Normal P-P plot of regression standardized residual.

The resulting model for unstandardized beta coefficients is:

$$\text{TOSLS} = -14.212 - 1.049(\text{EN}) + .936(\text{EP}) + .668(\text{GP}) + 1.772(\text{SC}) + .087(\text{SB}) + .041(\text{AS}) + .119(\text{SE}) + .020(\text{SA}) + .807(\text{GE}) + 2.511(\text{AI}) + 1.231(\text{AN}) + .858(\text{BK}) - .893(\text{HC}) + 1.065(\text{TE}) + .234(\text{WE})$$

EN = 0, 1	GP = 0.0 – 4.0	SC = 0 – (4.5)	AS = 0 - 114
EP = 0, 1	SB = 0 – 100	SE = 0 – 100	GE = 0, 1
	AI, AN, BK, HC, TE, WE = 0, 1		TOSLS = 0 – 28

Summary

This chapter presented data analysis methods to answer the research question. The data set was examined for outliers, normality, and homogeneity of variance prior to analysis. Interquartile ranges were calculated demonstrating that all TOSLS scores fell within the expected pattern for a normal distribution. A Kolmogorov-Smirnov test verified the normality of the population, $p = .000$, while Levene's Tests verified the homogeneity of variance for Gender $F(1, 374) = 0.260$, $p = .610$, Ethnicity $F(5, 370) = 1.036$, $p = .396$, economic need $F(1, 374) = 0.353$, $p = .553$, and English Proficiency $F(1, 374) = 0.036$, $p = .849$.

Variables were examined for significant differences in TOSLS scores. Categorical variables were assessed by *t*-test or ANOVA respectively while continuous variables were examined using a Pearson correlation. As illustrated in Table 27, all predictor variables, with the exception English Proficiency produced statistically significant results ($p < .05$).

Table 27

Summary of TOSLS Scores and Predictive Factors

Variable	Statistical Test	Value	Significance p
Gender	t-Test	-2.178	.030
Ethnicity	ANOVA	2.427	.035
English Proficiency	t-Test	-.384	.701
Economic Need	t-Test	3.722	.000
Science Courses Completed	Pearson Correlation	.202	.000
STAAR Biology	Pearson Correlation	.146	.005
STAAR ELA 1	Pearson Correlation	.498	.000
STAAR Algebra 1	Pearson Correlation	.434	.000
Grade Point Average	Pearson Correlation	.467	.000
BRAINS Total	Pearson Correlation	.385	.000

A hierarchal, multiple regression model was generated using all variables collected in the study. Predictive factors were categorized to account for the variations in TOSLS scores with distal factors accounting for 21.8%, opportunity factors for 6.5%, propensity factors for 15.3%, and demographic factors for 1.4% of variance. The model developed accounted for 45% of variance in TOSLS scores in totality.

Model diagnostics were conducted to determine the feasibility of the multiple regression model. An ANOVA demonstrated the model predicted TOSLS scores better than chance $F(7,368) = 43.111$, $p = .000$. Individual scores were examined using case wise residuals with

2.92% of scores falling outside two standard deviations of the model and a single case falling within three standard deviations consistent with a normalized population. The model was found to significantly predict scores on the TOSLS.

Chapter 5

DISCUSSION

Science literacy is the established goal of science education with little student progress from the inception of the term in 1958 to today. Science literacy has been incorporated into state and national learning standards with the development of various curricula to support it. Data for this study was collected from high school students to generate a predictive model of science literacy using the Test of Science Literacy Skills (TOSLS; Gormally et al., 2012) and to quantify variables that impact science literacy. Variables previously shown to impact science literacy including gender (Beaton & Chromy, 2010), ethnicity (Campbell et al., 1997), economic need (Kohlhaas et al., 2010), English proficiency (Grigg et al., 2006), science courses completed (Impey, 2017), Biology content knowledge (Nuhfer et al., 2016), English content knowledge (Tong et al., 2014), Algebra 1 content knowledge (Wang & Ma, 2016), grade point average (Impey, 2011) and attitude toward science (Newell et al., 2015) were included in the model. These variables were entered into a hierarchal, multiple regression model consistent with the distal, opportunity, propensity, and demographic model. The purpose of this chapter is to interpret key findings, discuss the implications of those findings, and suggest areas of future research.

Summary of Key Findings

The present study sought to determine the degree to which variation in demographic, distal, propensity, and opportunity factors predict the science literacy of high school students. Variables were classified as distal variables capable of impacting science literacy over greater duration and included Economic Need, English Proficiency, and Grade Point Average. Propensity variables were academic variables that are linked to science literacy performance.

The number of science courses completed was the only opportunity variable included in this study. Demographic factors collected for inclusion were Gender and Ethnicity.

Distal Factors

Distal factors, Economic Need, English Proficiency, and Grade Point Average have the potential to impact opportunity and propensity factors. Economic Need was defined as participation in the Free and Reduced Price (FARP) program at any point the student was enrolled in public schools due to the long-term effects as a distal variable. Students identified as in economic need (M = 12.13, SD = 4.900) score significantly lower on the TOSLS (p = .000) than the rest of the population (M = 14.50, SD = 5.008). The TOSLS of students classified as proficient in English (M = 13.99, SD = 5.091) were lower than those of students who had limited English proficiency students (M = 14.38, SD = 4.883) although the difference was not significant p = .701. GPA demonstrated a moderate, positive correlation $r(376)$ = .467, p = .000 with mean TOSLS scores.

A multiple regression model using only distal factors and TOSLS score was generated. The generated model demonstrated a positive, moderate, and significant relationship between distal variables and TOSLS scores that could account for 22% of the variance in science literacy. The impact of distal factors on science literacy did not rise to the level reported by Byrnes and Miller (2007), which accounted for a minimum of 28.8% of variance.

Opportunity Factors

The only opportunity factor that impacted students' ability to access the skills measured was Number of Science Courses Completed. The Number of Science Courses completed was separated into two categories. The first category included career and technology education (CTE) courses connected to a scientific, engineering, or medical career path while the second category included traditional science courses. The number of science courses completed had significant impact on the TOSLS scores (p = .000) with a weak, positive correlation between r(376) = .202. CTE courses did not demonstrate a significant correlation with TOSLS scores (r[376] = .067, p = .198) when examined independently. Science courses and CTE courses were combined into a single factor and correlated with TOSLS scores r(376) = .164, p = .001. Traditional science courses were a better predictor of science literacy when excluding CTE courses. CTE courses focus on a specific subset of knowledge rather than the principles and experimental methods behind the knowledge. These findings are consistent with those of Impey (2017), who found that that additional science courses result in greater science literacy. These findings also deviate from previous studies (National Academy of Engineering, 2006; Nuhfer et al., 2016) that found that the number of science courses had negligible impact on science literacy.

The regression model generated for opportunity factors and science literacy resulted in a weak, positive, significant correlation. Opportunity factors accounted for 3.8% of the variance in science literacy. Opportunity impact was lower than expected due to the inclusion of a single variable. Similar studies included student perceptions of the school, teacher, and learning opportunities outside of school.

Propensity Factors

Propensity factors reflect the ability or willingness to learn content and included measures of attitude toward science, biology content knowledge, algebra content knowledge, and English language arts content knowledge. Biology content knowledge, STAAR Biology, exhibited a positive, weak, and significant correlation with science literacy (r[376] = .146, p = .005). English Language Arts content knowledge, STAAR English Language Arts, had a moderate positive correlation with TOSLS performance r(376) = .498, p = .000. Algebraic content knowledge was correlated in a positive, moderate, and significant manner $r(376) = .434$, $p = .000$. The BRAINS assessment was positively and moderately correlated with mean TOSLS performance Attitude toward Science ($r[376] = .385$, $p = .000$).

Propensity factors, STAAR Biology, STAAR Algebra, STAAR English Language Arts, and BRAINS assessment were included in a multiple regression model with science literacy. Propensity factors exhibited a moderate, positive, and significant relationship with science literacy accounting for 35.5% of the variance.

Demographic Factors

Demographic factors were assessed for their impact on science literacy independent of other factors. Mean performance on the TOSLS was significant ($p = .030$) higher for male students ($M = 14.67$, $SD = 5.02$) when compared to female students ($M = 13.53$, $SD = 5.08$). Mean performance on the TOSLS test varied significantly between ethnic groups $F(5,370) = 2.427$, $p = .035$. Further exploration determined that the only significant differences between groups was that Asian students performed significantly better ($M = 17.31$, $SD = 4.771$) than Hispanic students ($M = 12.94$, $SD = 4.503$).

A multiple regression model was generated using only the demographic factors of gender and ethnicity with science literacy. The demographic variable model showed a very weak, positive correlation between demographic variables and science literacy. Changes in demographic factors could only account for 2.4% of the variance in TOSLS. This is consistent with the distal, propensity, opportunity model (Byrnes & Miller, 2007) that provides guidance on how to close achievement gaps.

Model

A hierarchal multiple regression model was generated by entering distal factors followed by opportunity, propensity and demographic factors, in order. All factors collected for the study were included in the model. The model for science literacy was moderately, positively, and significantly correlated with science literacy accounting for 44.5% of the variance in TOSLS. The model approaches a strong correlation with science literacy (r = 0.683, adjusted R^2 = .445).

Interpretation of Key Findings

The categorical variables of Gender, Ethnicity, Grade Point Average and Economic Need demonstrated significant differences in means between categories. The findings are consistent with previous literature and adds to the importance of considering Gender, Ethnicity, and Economic Need when designing curriculum and interventions in science literacy.

Economic Need was hypothesized to have the greatest impact on science literacy because of the distal nature of impacting propensity and opportunity variables. Free and Reduced Price (FARP) lunch programs, an indicator of economic need, provide nutritional assistance but do not provide the educational support necessary to overcome the opportunity and propensity gaps. Students with economic need have engaged with a greater number of science courses, but may lack the means to engage with science outside of school through museums, media, or parents. Students in the FARP program had lower performance in all propensity variables including Attitude Toward Science, indicating a need for early school intervention across multiple subjects. Intervention programs for students in economic need should shift from a deficit model to engaging students with opportunities outside the classroom to mirror the advantages of those with greater economic means. A focus on more than physiological needs may benefit attitude toward science as well as science literacy.

The gender gap narrowed with increased opportunity for women to engage with science. Initial analysis demonstrated significantly higher performance for male students consistent with previous literature. The gender difference dropped to 0.5% of variance in science literacy for this study when distal, opportunity, and propensity factors were accounted. Additionally, the inclusion of gender in the hierarchal model did not significantly change the predictability of the model. The initial difference in performance by gender can largely be explained with variance in the number of Science Courses Completed or less opportunity to engage with science literacy. Female students scored lower on the BRAINS assessment of Attitude Toward Science, impacting propensity for science literacy. Efforts to actively recruit and support women in science should be supported to encourage female participation in science courses and provide role models for future generations.

The number of participants identified as American Indian, Asian, Black, and Two or More races deviated from state and national averages and limited the generalizability of the findings regarding the relationship between ethnicity and science literacy. White (n = 251) and Hispanic (n = 66) populations are the only ethnicities that exceeded the rule of thumb established by the Central Limit Theorem (n = 30), which limited the significance of the findings. Preliminary analysis revealed a difference between groups; subsequent examination indicated that the only significant difference in scores was between Hispanic and Asian students. Ethnicity accounted for .6% of variance in science literacy according to the model when distal, opportunity, and propensity variables were factored. The discrepancy of higher mean performance for Asian students on all propensity factors indicates an unexamined distal factor is impacting science literacy performance.

The findings of this study deviated from the literature in the impact of English proficiency on science literacy. Contrary to findings in previous studies, in this study, students with limited English Proficiency had a higher mean performance on science literacy although this finding did not reach significance possibly due to the low number of participants (n = 26). The definition of limited English Proficiency may be the source of discrepancy. Participants were classified as limited English proficiency if they ever received TELPAS support or assessment including students exiting the program in elementary school. This may have led to a discrepancy in science literacy performance between Asian students (M = 17.8) and Hispanic students (M = 13.5) that resulted in students with limited English Proficiency (M = 14.38) outperforming students with English as their native language (M = 13.99). The limited sample size and problems with the definition limit the findings of English Proficiency on science literacy.

The Number of Science Courses completed represented the only opportunity variable accounted for in this study. A student completing one additional science course on average is predicted to score 1.8 more correct answers on the TOSLS. This represented 6.5% of the variance in science literacy according to the model. Students exceeding four science credits, which is required for most common graduation plans, scored higher on all propensity variables, which indicates that an unexamined distal factor may influence the willingness to take additional science courses.

The content knowledge propensity variables correlated with science literacy performance, which is consistent with previous studies. Content knowledge represented 13.2% of the variance in science literacy performance with Biology content knowledge accounting for 9.7% of the total. English Language Arts content knowledge as measured by the STAAR English Language

Arts had a stronger correlation with the science literacy than Algebra. Although it only accounted for 3.1% of the variance, this was counterintuitive. Science achievement has been historically linked to math achievement, but Algebra scores did not significantly impact the predictability of the model. Algebra scores represent only .4% of the variance in science literacy, the lowest of any variable included. This provides support for cross-curricular integration of Science and English classes to support literacy gains in both courses.

Attitude Toward Science, as measured by the BRAINS assessment, was positively correlated with science literacy, which is consistent with findings from previous studies. The final propensity variable significantly impacted the predictability of the model for science literacy and accounted for 2.3% of the variance in science literacy scores. Attitude Toward Science had substantial differences between ethnicities with Asian students reporting more positive attitudes toward science and received the higher average scores for science literacy. Grade Point Average (GPA) resulted in the greatest variance in science literacy. GPA was responsible for 18.6% of the variance in science literacy, which is consistent with its classification as an affective, distal variable. GPA did not significantly add to the predictability of the model even when accounting for the greatest variance. GPA includes cognitive and affective factors that may be manifest in the model with the inclusion of content knowledge, measures of attitude, and coursework.

The model generated in this study was moderately, positively, and significantly correlated with science literacy. The factors included account for 44% of the variance in TOSLS scores and can be used to target specific subskills for intervention in hopes of improving science literacy. The predictive model for science literacy with unstandardized coefficients and ranges for variables is shown below:

$$TOSLS = -14.212 - 1.049(EN) + .936(EP) + .668(GP) + 1.772(SC) + .087(SB) + .041(AS) + .119(SE) + .020(SA) + .807(GE) + 2.511(AI) + 1.231(AN) + .858(BK) - .893(HC) + 1.065(TE) + .234(WE)$$

EN = 0, 1	GP = 0.0 – 4.0	SC = 0 – (4.5)	AS = 0 - 114
EP = 0, 1	SB = 0 – 100	SE = 0 – 100	GE = 0, 1
AI, AN, BK, HC, TE, WE = 0, 1		TOSLS = 0 – 28	

The impact of each variable that can be affected by schools was examined to prioritize intervention to improve science literacy as measured by the TOSLS. The maximum value for each factor was entered into the model and compared with the average for that factor. The difference between the maximum value and the value according to the data set was computed to prioritize topics or intervention. The number of science courses a student completes has the greatest potential impact on the development of science literacy at this location, followed by cross-curricular development between science and English teachers (Adams & Pegg, 2012). An additional intervention could focus on promoting students' positive attitude toward science. The impact of each variable on TOSLS score is illustrated in Table 28.

The impact was determined by examining the difference in the maximum and mean performance in the study and multiplying by the Beta coefficient to examine the potential impact of targeting that variable with intervention. The resulting impact uses the model to predict the potential increase in the number of TOSLS questions correct. Science Courses Completed has the potential to increase TOSLS performance by 3.69 but would require significant policy changes and the allocation of additional resources. English Language content knowledge, as measured by the STAAR ELA 1, can increase TOSLS performance by 2.28 questions (8%), followed by Attitude Toward Science by 1.69 questions (6%). The potential of these two variables has implications for research and practice.

Table 28

Potential Impact of Predictor Variables on TOSLS Scores

Variable	Min	Max	TOSLS Min	TOSLS Max	Average	TOSLS Avg	Impact
STAAR Alg 1	0	100.	0	2	80.48	1.61	0.39
Grade Point Average	0	4.	0	2.672	3.389	2.26	0.41
STAAR Biology	0	100.	0	8.7	81.9	7.13	1.57
BRAINS	0	114.	0	4.67	72.69	2.98	1.69
STAAR ELA 1	0	100.	0	11.9	80.85	9.62	2.28
Science Courses Completed	0	4.5	0	7.97	2.42	4.28	3.69

Implications for Research and Practice

The statistically significant difference between historically underserved populations in science literacy should be addressed prior to their manifestation in high school. The performance gap between male and female students has narrowed (Campbell et al., 1997) but remains despite decades of intervention at the elementary and secondary levels. The lack of significant differences in TOSLS scores for ethnic groups, and minimal impact on the model, indicates supports the development of interventions focused on distal variables including language acquisition. The emphasis on high stakes assessments limits students' exposure to information that later impacts students their educational career. The development of a standards-based curriculum with an emphasis on content area literacy can reduce the impact of these categorical variables. Students' limited exposure to content knowledge directly impacts their long-term acquisition of knowledge and science literacy skills necessary for success in the modern economy.

The instrumentation used in this study provides important information that can be used to measure science literacy and attitude toward science among high school students. The TOSLS was originally developed for use with undergraduate students (Gormally et.al, 2012) but the instructional expectations, content, and results demonstrated efficacy with high school students. Expanding the use of the TOSLS would provide low-stakes information on progress toward a scientifically literate population. The BRAINS assessment was originally developed and validated (Summers & Abd-El-Khalick, 2018) with 5th–10th grade students but demonstrated consistent results with 11th and 12th grade students. The importance of attitude toward science in science literacy and achievement justifies the use of the BRAINS as a course evaluation tool. The electronic distribution and sensitivity of these assessments could provide significant information on within-semester or year interventions as well as longitudinal information that impacts long-term gains.

The substitution of career and technology education courses for traditional science courses should be reversed and these courses should be used as general electives. The expansion of funding for CTE courses, including additional funds, has become widespread in high schools, which aim to prepare students for college, a career, or the military. A number of these courses can be counted as substitutes for traditional science courses including Anatomy and Physiology or Forensic Science. In this study, there was no appreciable correlation between these courses and science literacy. The inclusion of CTE courses as science courses negatively impacted the correlation between the number of science courses taken and science literacy. The negative impact on science literacy may be stronger than the career exposure through CTE courses.

Next Generation Science Standards (NGSS) should be adopted to support cross-curricular integration. NGSS provide a K-12 roadmap to develop a scientifically literate population using content and cross-cutting themes with targeted measurable outcomes at each stage of science literacy development. The integration of literacy and mathematical concepts into curriculum promotes cross-curricular planning and lesson development between teachers. The universal adoption of the NGSS would allow for a standardization of opportunity with early identification for students in need of intervention.

Instructional design should support a partnership between English Language Arts courses and Science courses. English Language Arts had a greater potential impact on science literacy than Science content knowledge or Algebra content knowledge. The focus of science intervention may include content acquisition and math skill development but should add the acquisition of general literacy skills to maximize gains in science literacy. Science intervention that includes reading and writing as a corollary to scientific inquiry can support the instructional objectives of both courses.

The development of a positive attitude toward science is directly correlated with higher student literacy performance. Learning standards are currently packed with content providing little opportunity for teachers and students to develop an appreciation for science and the impact it has on their daily life. State lawmakers should include science appreciation directly into learning standards to support the development of positive attitudes, increase the number of students enrolled in science coursework, and support the science literacy of all students.

The purpose of science education should include content acquisition, development of science literacy skills, and appreciation of science as cornerstone of society. The pendulum has continued to oscillate between content acquisition and skills for over a century without students reaching an expected levels of performance. Attitude toward science is correlated with the development of science literacy skills necessary to engage with the content. Engagement with the content leads to acquisition of knowledge, which can improve attitude. Content acquisition, science literacy skills, and appreciation of science are not oppositional approaches but instead represent reinforcement of approaches that conflict only when one is emphasized more than the others.

Limitations

The findings of this study should be interpreted after acknowledging the limitations of the study. Participants self-selected from a sample of convenience at a single high school in the southwestern United States. The demographic makeup of the student population of the campus significantly differed from state and national averages in terms of the percentage of minority, economically disadvantaged, and English language learner students. The lack of congruence between the participants and the larger population limit the generalizability of the findings. The limited participant size (n = 376) and number of variable categories did not result in an adequate sample size for English proficiency.

This study relied heavily on the accuracy of existing educational records. The review of educational records, particularly STAAR scores, for transient students relies on the manual input of information that is subject to human error. Examination of transcripts revealed six transcripts in which credits did not correspond to grade level classification. Those participants were removed from dataset prior to analysis but illustrate a potential source of error.

This study utilized self-report assessments to measure attitude toward science and to assess science literacy. Self-report measures could result in a bias of response based on the desired response expected by the teacher, researcher, or peer group. The researcher has an established relationship with participants in the school that could influence answers even with an emphasis on anonymity prior to assessment.

Recommendations for Future Research

The present study addressed a gap in the literature by contributing to the evidence supporting and quantifying the impact of the predictor variables on science literacy. Future research should seek to extend the findings of this model to other participant populations that correspond to state and national demographic trends. The findings of this study have limited generalizability as the participants were from a single suburban high school. The model should be assessed with schools that mirror the larger population to support targeted intervention based on the findings.

Future research should revisit English proficiency as a predictor variable. Nearly 20% of the U.S. student population are identified as English language learners. The limited sample size (n = 26) and definition of English proficiency in the present study contradicted previous research (Grigg et al., 2006; Nuhfer et.al, 2016) that English proficiency resulted in significantly lower scores in science literacy. English proficiency was included as a binary, categorical variable based on the impact of language acquisition on literacy over time. Future studies should seek to measure English proficiency as a continuous, interval variable based on current status in an English proficiency program.

The instruments used in this study need further research to assess the relationship between the subscales. The design of this study assumed equal impact of the five subscales of the BRAINS instrument on science literacy. Further research could assess if any of the subscales have a disproportionate impact on science literacy as a whole or for particular skills. The directionality of science literacy as an outcome variable should be assessed in future research. Prior achievement has the ability to impact achievement possibly limiting the impact of attitude interventions to support science literacy. The TOSLS is moderately and positively correlated with statewide assessments used for school accountability. The TOSLS and BRAINS are relatively new instruments. Future research should explore the directionality of the TOSLS test in correlation with state and national assessments.

Conclusion

Findings from this research suggest that science literacy is a multimodal construct that depends on factors not directly linked to science content acquisition. The distal, opportunity, and propensity model accounted for 44.5% of the variance in science literacy. This study supports findings of previous research and extends those findings to high school students. Quantifying the impact of each variable allows for a targeted approach to developing intervention. Cross-curricular conversations between English Language Arts teachers and science teachers can impact the performance of students in science literacy more than that of science and algebra teachers.

The findings of this can help educators develop targeted interventions to support a more scientifically literate population. A curriculum that focuses on improving student attitudes toward science and the development of cross-curricular connections has more impact on science literacy than increasing the number of career and technology education courses student take. The increased sophistication and pace of change in society necessitates the development of students' science literacy skills so that they can adapt to those changes. Providing students with the skills necessary for success in the current economic climate should be the goal for science education.

REFERENCES

Aberšek, M., & Aberšek, B. (2013). A reading curriculum for the homo zapiens generation: New challenges, new goals. *Journal of Baltic Science Education*, *12*(1), 92–106.

Adams, A. E., & Pegg, J. (2012). Teachers' enactment of content literacy strategies in secondary science and mathematics classes. *Journal of Adolescent and Adult Literacy*, *56*(2), 151–161. doi:10.1002/JAAL.00116

Ajzen, I. (1985). From intentions to actions: A theory of planned behavior. In J. Kuhl & J. Beckman (Eds.), *Action-control: From cognition to behavior*. Heidelberg, Germany: Springer.

Ajzen, I. (2011). The theory of planned behavior: Reactions and reflections. *Psychology and Health*, *26*(9), 1113–1127. doi:10.1080/08870446.2011.613995

Ajzen, I. (2012). Martin Fishbein's legacy: The reasoned action approach. *The Annals of the American Academy of Political and Social Science*, *640*(1), 11–27. doi:10.1177/0002716211423363

Allchin, D., Andersen, H. M., & Nielsen, K. (2014). Complementary approaches to teaching nature of science: Integrating student inquiry, historical cases, and contemporary cases in classroom practice. *Science Education*, *98*(3), 461–486. doi:10.1002/sce.21111

Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, *31*(2), 211–236. doi:10.3386/w23089

Allum, N., Besley, J., Gomez, L., & Brunton-Smith, I. (2008). Disparities in science literacy. *Science*, *360*(6391), 861–862. doi:10.1126/science.aar8480

American Association for the Advancement of Science. (1990). *Project 2061: Science for all Americans.* New York, NY: Oxford University Press.

American College Testing (2015). *The condition of college and career readiness*. (pp. 1-40). Retrieved from https://www.act.org/readiness/2015

American Library Association. (2000). *Information literacy competency standards for higher education*. Chicago, IL: Association of College & Research Libraries.

Atwater, M. M., Wiggins, J., & Gardner, C. M. (1995). A study of urban middle school students with high and low attitudes toward science. *Journal of Research in Science Teaching*, *32*(6), 665–677.

Auerbach, A. J., & Schussler, E. E. (2017). Curriculum alignment with vision and change improves student scientific literacy. *Cell Biology Education*, *16*(2), 1–9. doi:10.1187/cbe.16-04-0160

Bailey, P. (1998). Conceptions of scientific literacy: Making sense of a proposed national science curriculum framework. *Alberta Science Education Journal*, *30*(2), 52–59.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*(2), 191.

Beaton, A. E., & Chromy, J. R. (2010). *National assessment of educational progress trends: Main NAEP vs. long-term trend*. San Mateo, CA: American Institutes for Research.

Beaton, A. E., Martin, M. O., Mullis, I. V. S., Gonzalez, E. I., Smith, T. A., & Kelly, D. L. (1998). Create news; Consortium for research on educational accountability and teacher evaluation; Mathematics and science achievement in the middle school years: An international perspective. *Journal of Personnel Evaluation in Education*, *12*(1), 69–73. doi:10.1023/A:1007968617286

Biancarosa, G., & Snow, C. (2004). *Reading* next. New York, NY: Carnegie Corporation. Retrieved from https://www.carnegie.org/publications/reading-next/

Blank, R. K. (2013). Science instructional time is declining in elementary schools: What are the implications for student achievement and closing the gap? *Science Education*, *97*(6), 830–847. doi:10.1002/sce.21078

Bound, J., & Turner, S. (2002). Going to war and going to college: Did world war II and the G.I. bill increase educational attainment for returning veterans? *Journal of Labor Economics*, *20*(4), 784–815. doi:10.1086/342012

Burghardt, M. D., Lauckhardt, J., Kennedy, M., Hecht, D., & McHugh, L. (2015). The effects of a mathematics infusion curriculum on middle school student mathematics achievement. *School Science and Mathematics*, *115*(5), 204–215.

Bybee, R. W. (1997). *Achieving scientific literacy: From purposes to practices*. Portsmouth, NH: Heinemann.

Bybee, R. W., & McInerney, J. D. (1995). *Redesigning the science curriculum*. Colorado Springs, CO: Biological Sciences Curriculum Study. Retrieved from https://files.eric.ed.gov/fulltext/ED433179.pdf

Campbell, J. R., Hombo, C. M., & Mazzeo, J. (2000). NAEP 1999 trends in academic progress: Three decades of student performance. *Education Statistics Quarterly*, *2*(4), 31.

Campbell, J. R., Voelkl, K. E., & Donahue, P. L. (1997). *NAEP 1996 Trends in academic progress*. (NCES Report No. 97-985). Retrieved from U.S. Department of Education website: https://files.eric.ed.gov/fulltext/ED424301.pdf

Caskie, G. I. L., Sutton, M. C., & Eckhardt, A. G. (2014). Accuracy of self-reported college GPA: Gender-moderated differences by achievement level and academic self-efficacy. *Journal of College Student Development*. *55*(4), 385– 390. doi:10.1353/csd.2014.0038

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.

Chandler, J. R. (2017). [Relationship between science literacy, motivation, and achievement]. Unpublished raw data.

Chapline, G. F. (1991, March). Megascience: Large projects yield little benefit. *USA Today*, pp. 52–54.

Coke, P. (2014). *Determining the alignment between what teachers are expected to teach, what they know, and how they assess scientific literacy* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3624327)

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues*. Boston, MA: Houghton Mifflin.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, *78*(1), 98.

Crede, M., & Phillips, L. A. (2011). A meta-analytic review of the motivated strategies for learning questionnaire. *Learning and Individual Differences*, *21*(4), 337–346. doi:10.1016/j.lindif.2011.03.002

Cruse, C., & Powers, D. (2004). *Estimating school district poverty with free and reduced-price lunch data*. U.S. Census Bureau, Small Area Estimates Branch. Retrieved from https://www.census.gov/content/dam/Census/library/working-papers/2006/demo/crusepowers2006asa.pdf

Czerniak, C. M., Weber, W. B., Jr., Sandmann, A., & Ahern, J. (1999). A literature review of science and mathematics integration. *School Science and Mathematics*, *99*(8), 421–430.

Deboer, G. E. (2000). Scientific literacy another look. *Journal of Research in Science Teaching*, *37*(6), 582–601.

Dee, T., & Jacob, B. (2010). Evaluating NCLB. *Education Next*, *10*(3), 54–61.

Department of the Interior, Bureau of Education. (1918). *Cardinal principles of secondary education: A report of the commission on the reorganization of secondary education, appointed by the national education association* (Bulletin No. 3). Retrieved from https://eric.ed.gov/?id=ED541063

Dredze, M., Broniatowski, D. A., Smith, M. C., & Hilyard, K. M. (2016). Understanding vaccine refusal: Why we need social media now. *American Journal of Preventive Medicine, 50*(4), 550–552. doi:10.1016/j.amepre.2015.10.002

Duke, N. K. (2000). 3.6 minutes per day: The scarcity of informational texts in first grade. *Reading Research Quarterly*, *35*(2), 202–224.

Duncan, T. G., & McKeachie, W. J. (2005). The making of the motivated strategies for learning questionnaire. *Educational Psychologist*, *40*(2), 117–128. doi:10.1207/s15326985ep4002_6

Durant, J. (1994). What is scientific literacy? *European Review*, *2*(1), 83–89.

Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist*, *41*(10), 1040–1048. doi:10.1037/0003-066X.41.10.1040

Elementary and Secondary Education Act, Pub. L. No. 89-10, § 79, (1965).

Every Student Succeeds Act of 2015, Pub. L. No. 114-95, § 114, (2015).

Faulkner, S. F. (2012). Science literacy: Exploring middle-level science curriculum structure and student achievement. *NERA Conference Proceedings 2012*, Paper 18. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.1005.7838&rep=rep1&type=pdf

Feinstein, N. (2011). Salvaging science literacy. *Science Education*, *95*(1), 168–185. doi:10.1002/sce.20414

Field, A. P. (2013). *Discovering statistics using IBM SPSS statistics: And sex and drugs and rock 'n' roll* (4th ed.). Los Angeles, CA. Sage.

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention, and behavior: An introduction to theory and research.* Reading, MA: Addison-Wesley.

Fishbein, M., & Ajzen, I. (2010). *Predicting and changing behavior: The reasoned action Approach*. New York, NY: Psychology Press.

Fleischman, H. L., Hopstock, P. J., Pelczar, M. P., Shelley, B. E., & National Center for Education Statistics. (2010). *Highlights from PISA 2009: Performance of U.S. 15-year-old students in reading, mathematics, and science literacy in an international context*. Washington, DC: U.S. Department of Education.

Foster, J., & Shiel-Rolle, N. (2011). Building scientific literacy through summer science camps: A strategy for design, implementation and assessment. *Science Education International*, *22*(2), 85–98. Retrieved from http://www.icaseonline.net/sei/june2011/p1.pdf\

Funk, C., Rainie, L., & Page, D. (2015). Public and scientists' views on science and society. *PEW Research Center.* Retrieved from https://www.pewresearch.org/science/2015/01/29/public-and-scientists-views-on-science-and-society/

Gabel, L. L. (1976). *The development of a model to determine perceptions of scientific literacy* (Doctoral dissertation, Ohio State University). Retrieved from https://etd.ohiolink.edu/!etd.send_file?accession=osu1487002706775026&disposition=attachment

Gardner, P. L. (1995). Measuring attitudes to science. *Research in International Journal of Science Education*, *26*(1), 141–169. doi:10.1007/BF02357402

Garner-O'Neale, L., & Ogunkola, B. (2015). Effects of interest in science, study habits, sex and level of study on the nature of science literacy level of undergraduate chemistry students of the university of the West Indies, Barbados. *Journal of Educational and Social Research*, *5*(2), 267–274. doi:10.5901/jesr.2015.v5n2p267

George, R. (2006). A cross-domain analysis of change in students' attitudes toward science and attitudes about the utility of science. *International Journal of Science Education*, *28*(6), 571–589. doi:10.1080/09500690500338755

Georgia Department of Education. (2016). *Science Georgia standards of excellence: Chemistry standards*. Retrieved from https://www.georgiastandards.org/Georgia-Standards/Documents/Science-Chemistry-Georgia-Standards.pdf

Germann, P. J. (1988). Development of the attitude toward science in school assessment and its use to investigate the relationship between science achievement and attitude toward science in school. *Journal of Research in Science Teaching*, *25*(8), 689.

Gormally, C., 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), 364–377. doi:10.1187/cbe.12-03-0026

Graff, Z. A. (2016). *Focal points: Affecting undergraduates' scientific literacy through a three-skill intervention* (Master's thesis). Available from ProQuest Dissertations and Theses databases. (UMI No. 10128036)

Green, J., Martin, A. J., & Marsh, H. W. (2007). Motivation and engagement in English, mathematics and science high school subjects: Towards an understanding of multidimensional domain specificity. *Learning and Individual Differences*, *17*(3), 269–279. doi:10.1016/j.lindif.2006.12.003

Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test III: Meta-analysis of predictive validity. *Journal of Personality and Social Psychology, 97*(1), 17–41.

Grigg, W., Lauko, M., & Brockway, D. (2006). *The Nation's Report Card: Science 2005* (NCES 2006-466). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Guzzetti, B. J., & Bang, E. (2010). The influence of literacy-based science instruction on adolescents' interest, participation, and achievement in science. *Literacy Research and Instruction*, *50*(1), 44–67. doi:10.1080/19388070903447774

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, *66*(1), 64–74. doi:10.1119/1.18809

Hayes, B. C., & Tariq, V. N. (2001). Gender differences in scientific knowledge and attitudes toward science: A reply to a reply. *Museum*, *10*(1), 431–433. doi:10.1088/0963-6625/9/4/306

Hazen, R., & Trefil, J. (1991). *Science matters: Achieving science literacy*. New York, NY: Anchor.

Helgeson, S., Blosser, P. E., & Howe, R. W. (1977). *The status of pre-college science, mathematics, and social science education: 1955-1975* (National Science Foundation Report No. 7620627). Retrieved from https://files.eric.ed.gov/fulltext/ED153876.pdf

Hellmuth, R. (2014). *Predicting science literacy and science appreciation* (Honor's thesis, University of Central Florida). Retrieved from https://stars.library.ucf.edu/cgi/viewcontent.cgi?article=2667&context=honorstheses1990-2015

Hertzfeld, H. R. (2002). Measuring the economic returns from successful NASA life sciences technology transfers. *The Journal of Technology Transfer*, *27*(4), 311–320. doi:/10.1023/A:1020207506064

Holbrook, J., & Rannikmae, M. (2007). The nature of science education for enhancing scientific literacy. *International Journal of Science Education*, *29*(11), 1347–1362. doi:10.1080/09500690601007549

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

Hurd, P. D. (1970). Scientific enlightenment for an age of science. *The Science Teacher*, *37*(1), 13–15.

İlker, G. E., Arslan, Y., & Demirhan, G. (2014). A validity and reliability study of the motivated strategies for learning questionnaire. *Educational Sciences: Theory & Practice*, *14*(3), 829–834. doi:10.12738/estp.2014.3.1871

Impey, C., Buxner, S., Antonellis, J., Johnson, E., & King, C. (2011). A twenty-year survey of science literacy among college undergraduates. *Journal of College Science Teaching*, *40*(4), 31–37.

Impey, C., Formanek, M., Sanlyn, R. B., & Wenger, M. C. (2017). Twenty seven years of tracking undergraduate science knowledge and beliefs. *Electronic Journal of Science Education*, *21*(4), 41–64.

Jarman, R., & McClune, B. (2002) A survey of the use of newspapers in science instruction by secondary teachers in Northern Ireland. *International Journal of Science Education*, *24*(10), 997–1020. doi:10.1080/09500690210095311

Jones, L. R. (1992). *The 1990 science report card: NAEP's assessment of fourth, eighth, and twelfth graders*. Washington, DC: US Government Printing Office.

Kaiser, D. (2006). The physics of spin: Sputnik politics and American physicists in the 1950s. *Social Research*, *73*(4), 1225–1252.

Kena, G., Hussar, W., McFarland, J., de Brey, C., Musu-Gillette, L., Wang, X., & Ossolinski, M. (2016). *The condition of education 2016*. Washington, DC: United States Department of Education. Retrieved from https://nces.ed.gov/pubs2016/2016144.pdf

Kirsch, I. S., Jungeblut, A., Jenkins, L., & Kolstad, A. (2002). *Adult literacy in America: A first look at the results of the national adult literacy survey* (3rd ed.). Washington, DC: US Government Printing Office. Retrieved from https://nces.ed.gov/pubs93/93275.pdf

Klucevsek, K. M., & Brungard, A. B. (2016). Information literacy in science writing: How students find, identify, and use scientific literature. *International Journal of Science Education*, *38*(17), 2573–2595. doi:10.1080/09500693.2016.1253120

Kohlhaas, K., Lin, H., & Chu, K. (2010). Disaggregated outcomes of ethnicity, gender, and poverty on fifth grade science performance. *Research in Middle Level Education Online*, *33*(6), 1–12.

Kuncel, N. R., Credé, M., & Thomas, L. L. (2005). The validity of self-reported grade point averages, class ranks, and test scores: A meta-analysis and review of the literature. *Review of Educational Research*, *75*(1), 63–82. doi:10.3102/00346543075001063

Kutner, M., Greenberg, E., Jin, Y., Boyle, B., Hsu, Y. C., & Dunleavy, E. (2007). *Literacy in everyday life: Results from the 2003 National Assessment of Adult Literacy*. Washington, DC: U.S. Department of Education. Retrieved from https://nces.ed.gov/Pubs2007/2007480_1.pdf

Laugksch, R. C. (2000). Scientific literacy: A conceptual overview. *Science Education*, *84*(1), 71–94.

Layton, D., Jenkins, E., & Donnelly, J. (1994). *Scientific and technological literacy: Meanings and rationales: An annotated bibliography*. Leeds, England: University of Leads Centre for Studies in Science and Mathematics Education.

Lee, K. M., & Slate, J. R. (2014). Differences in advanced achievement outcomes for Texas students as a function of economic disadvantage. *Journal of Education Research*, *8*(3), 137–149.

Lemke, J. L. (1998). *Multiplying meaning: Visual and verbal semiotics in scientific text*. In J. R. Martin & R. Veel (Eds.), *Reading science* (pp. 87–113). London, England: Routledge.

Li, C. H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods*, *48*(3), 936–949.

Liu, X. (2009). Beyond science literacy: Science and the public. *International Journal of Environmental and Science Education*, *4*(3), 301–311.

Madden, T., Ellen, P., & Ajzen, I. (1992). A comparison of the theory of planned behavior and the theory of reasoned action. *Personality and Social Psychology Bulletin*, *18*(1), 3–9.

Maltese, A. V., Tai, R. H., & Fan, X. (2012). When is homework worth the time? Evaluating the association between homework and achievement in high school science and math. *The High School Journal*, *96*(1), 52–72. doi:10.1353/hsj.2012.0015

Manyika, J., & Roxburgh, C. (2011). The great transformer: The impact of the Internet on economic growth and prosperity. *McKinsey Global Institute*. Retrieved from https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/the-great-transformer#

Martin, M. O., Mullis, I. V. S., Foy, P. F., & Stanco, G. (2012). *TIMSS 2011 international results in science*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. Retrieved from http://timssandpirls.bc.edu/timss2011/international-results-science.html

Marx, R., & Harris, C. (2006). No Child Left Behind and science education: Opportunities, challenges, and risks. *The Elementary School Journal*, *106*(5), 467–478. doi:10.1086/505441

McFarland, J., Hussar, B., de Brey, C., Snyder, T., Wang, X., Wilkinson-Flicker, S., . . . Hinz, S. (2017). *The condition of education 2017*. Washington, DC: U.S. Department of Education. Retrieved from https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2017144

McQuaid, K. (2007). Sputnik reconsidered: Image and reality in the early space age. *Canadian Review of American Studies*, *37*(3), 371–401. doi: 10.3138/cras.37.3.371

Millar, R. (2002). Towards a science curriculum for public understanding. In S. Amos & R. Boohan (Eds.), *Teaching science in secondary schools: A reader* (pp. 112–128). London, England: Routledge.

Miller, J. D. (1983). Scientific literacy: A conceptual and empirical review. *Daedalus*, *112*(2), 29–48.

Miller, J. D. (1998). The measurement of civic scientific literacy. *Public Understanding of Science*, *7*(3), 203–223. doi:10.1088/0963-6625/7/3/001

Miller, J. D. (2012). What colleges and universities need to do to advance civic scientific literacy and preserve American democracy. *Liberal Education*, *98*(4), 28–33. Retrieved from https://www.aacu.org/liberaleducation/le-fa12/miller.cfm

Mohadjer, L., Kalton, G., Krenzke, T., Liu, B., Van de Kerckhove, W., Li, L., . . . White, S. (2009). *National assessment of adult literacy: Indirect county and state estimates of the percentage of adults at the lowest literacy level for 1992 and 2003* (Research Report No. 2009-482). Retrieved from http://eric.ed.gov/?id=ED503830

Mondak, J. J., & Ganache, D. (2004). Knowledge variables in cross-national social inquiry. *Social Science Quarterly*, *85*(3), 539–558. doi:10.1111/j.0038-4941.2004.00232.x

Moreno, N. P., Tharp, B. Z., Vogt, G., Newell, A. D., & Burnett, C. A. (2016). Preparing students for middle school through after-school STEM activities. *Journal of Science Education and Technology*, *25*(6), 889–897.

Moss, B. (2005). Making a case and a place for effective content area literacy instruction in the elementary grades. *The Reading Teacher*, *59*(1), 46–55. doi:10.1598/rt.59.1.5

Mullis, I. V. S., & Jenkins, L. B. (1988). *The science report card: Elements of risk and recovery. Trends and achievement based on the 1986 national assessment*. Washington, DC: Office of Educational Research and Improvement.

Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2016). *TIMSS 2015 international results in mathematics*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. Retrieved from http://timssandpirls.bc.edu/timss2015/international-results/

Murti, P. R., Aminah, N. S., & Harjana. (2018). The analysis of high school students' science literacy based on nature of science literacy test (NOSLiT). *Journal of Physics: Conference Series*, *1097*, 12003. doi:10.1088/1742-6596/1097/1/012003

Nasr, A. R., & Soltani, A. (2011). Attitude towards biology and its effects on students' achievement. *International Journal of Biology*, 3(4), 100–104.

National Academy of Sciences, National Academy of Engineering, and Institute of Medicine. 2007. *Rising above the gathering storm: Energizing and employing America for a brighter economic future*. Washington, DC: The National Academies Press. doi:10.17226/11463

National Assessment of Educational Progress. (2015a). Science assessment. The nation's report card. Retrieved from https://www.nationsreportcard.gov/science_2015/#?grade=4

National Assessment of Educational Progress. (2015b). Science assessment. The nation's report card. Retrieved from https://www.nationsreportcard.gov/science_2015/#gaps/chart_loc_1?grade=12

National Center for Education Statistics (2011). *The nation's report card: Science 2009*. Washington, DC: U.S. Government Printing Office.

National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform.* Washington, DC: U.S. Government Printing Office. Retrieved from https://eric.ed.gov/?id=ED226006

National Commission on Mathematics and Science Teaching for the 21st Century. (2000). *Before it's too late: A report to the nation from the national commission on mathematics and science teaching for the 21st century*. Washington, DC: U.S. Department of Education.

National Defense Education Act of 1958, Pub. L. No. 85-864. (1958).

National Research Council. (1996). *National science education standards*. Washington, DC: The National Academies Press. doi:10.17226/4962.

National Research Council. (2012). *A framework for K–12 science education: Practices, crosscutting concepts, and core Ideas*. Washington, DC: The National Academies Press. doi:10.17226/13165

National Research Council. (2013). *Next generation science standards: For states, by states*. Washington, DC: The National Academies Press. doi:10.17226/18290

National Research Council. (2014). *Literacy for science: Exploring the intersection of the next generation science standards and common core for ELA standards: A workshop summary*. Washington, DC: The National Academies Press. doi:10.17226/18803

National Science Board. (2020). Bachelor's degree holders among individuals 25–44 years old. *National science foundation*. Retrieved from https://ncses.nsf.gov/indicators/states/indicator/bachelors-degree-holders-per-25-44-year-olds

National Science Foundation. (2006). *America's pressing challenge: Building a stronger foundation*. Arlington, VA: National Science Foundation.

National Student Clearinghouse. (2015). *Post-secondary enrollment and progress: Vol. RHS*. Herndon, VA: National Student Clearinghouse.

Neidorf, T., Stephens, M., Lasseter, A., Gattis, K., Arora, A., Wang, Y., . . . Rahman, T. (2015). *A comparison between the NGSS and the NAEP frameworks in science, technology and engineering literacy, and mathematics*. Washington, DC: National Center for Education Statistics. Retrieved from https://nces.ed.gov/nationsreportcard/subject/science/pdf/ngss_naep_technical_report.pdf

Newell, A. D., Zientek, L. R., Tharp, B. Z., Vogt, G. L., & Moreno, N. P. (2015). Students' attitudes toward science as predictors of gains on student content knowledge: Benefits of an after-school program. *School Science and Mathematics*, *115*(5), 216–225.

Next Generation Science Standards Lead States. (2013). *Next generation science standards: For states, by states*. Washington, DC: The National Academies Press. doi:10.17226/18290

No Child Left Behind Act of 2001, Pub. L. No. 107-110, § 6319 (2002).

Noddings, N. (1992). *The challenge to care in schools: An alternative approach toeducation*. New York, NY: Columbia University Teachers College Press.

Noftle, E. E., & Robins, R. W. (2007). Personality predictors of academic outcomes: Big five correlates of GPA and SAT scores. *Journal of Personality and Social Psychology*, *93*(1), 116–130. doi:10.1037/0022-3514.93.1.116

Nuhfer, E. B., Cogan, C. B., Kloock, C., Wood, G. G., Goodman, A., Delgado, N. Z., & Wheeler, C. W. (2016). Using a concept inventory to assess the reasoning component of citizen-level science literacy: Results from a 17,000-student study. *Journal of Microbiology & Biology Education*, *17*(1), 143–155. doi:10.1128/jmbe.v17i1.1036

Olatoye, R. A. (2009). Students' test anxiety, motivation for examinations and science achievement in junior secondary schools in Ogun State, Nigeria. *International Journal of Psychology and Counselling*, *1*(10), 194–198.

O'Reilly, T., & McNamara, D. S. (2007). The impact of science knowledge, reading skill, and reading strategy knowledge on more traditional "high-stakes" measures of high school students' science achievement. *American Educational Research Journal, 44*(1), 161-196.

Organisation for Economic Co-Operation and Development. (2000). *Measuring student knowledge and skills: The PISA 2000 assessment of reading, mathematical, and scientific literacy*. Retrieved from http://www.oecd.org/education/school/programmeforinternationalstudentassessmentpisa/33692793.pdf

Organisation for Economic Co-Operation and Development. (2007). *PISA 2006: Science competencies for tomorrow's world: Vol. 1: Analysis*. doi:10.1787/9789264040014-en

Organisation for Economic Co-Operation and Development. (2016). *PISA 2015 results: Vol. I: Excellence and equity in education*. doi:10.1787/9789264266490-en

Organisation for Economic Co-Operation and Development. (2017). *PISA for development brief 10: How does PISA for Development measure scientific literacy?* Retrieved from https://www.oecd.org/pisa/pisa-for-development/10-How-PISA-D-measures-science-literacy.pdf

Organisation for Economic Co-Operation and Development. (2019). *Science performance: PISA indicator*. doi:10.1787/91952204-en

Osborne, J., Simon, S., & Collins, S. (2003). Attitudes towards science: A review of the literature and its implications. *International Journal of Science Education*, *25*(9), 1049–1079. doi:10.1080/0950069032000032199

Osborne, J. W., & Overbay, A. (2004). The power of outliers (and why researchers should always check for them). *Practical Assessment, Research & Evaluation*, *9*(6), 1–8.

O' Sullivan, C. Y., Lauko, M. A., Grigg, W. S., Qian, J., & Zhang, J. (2003). *The nation's report card: Science 2000*. Washington, DC: National Center for Education Statistics.

Owen, S. V., Toepperwein, M. A., Marshall, C. E., Lichtenstein, M. J., Blalock, C. L., Liu, Y., . . . Grimes, K. (2008). Finding pearls: Psychometric reevaluation of the Simpson–Troost Attitude Questionnaire (STAQ). *Science Education*, *92*(6), 1076–1095.

Peng, S. (1995). *Understanding racial-ethnic differences in secondary science and mathematics achievement*. Washington, DC: National Center for Education Statistics.

Pintrich, P. R., & DeGroot, E. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, *82*(1), 33–40. doi: 10.1037/0022-0663.82.1.33

Pintrich, P. R., Smith, D., Garcia, T., & McKeachie, W. (1993). Reliability and predictive validity of the motivated strategies for learning questionnaire (MSLQ). *Educational and Psychological Measurement*, *53*(3), 801–813. doi:10.1177/0013164493053003024

Rideout, V. J., Foehr, U. G., & Roberts, D. F. (2010). Generation M2: Media in the lives of 8 to 18-year-olds. *Kaiser family foundation*. Retrieved from www.kff.org/entmedia/mh012010pkg.cfm

Roberts, D. A. (2007). Scientific literacy/Science literacy. In S. K. Abell & N. G. Lederman (Eds.), *Handbook of research in science education* (pp. 729–779). London, England: Routledge.

Roblin, N. P., Schunn, C., & McKenney, S. (2018). What are critical features of science curriculum materials that impact student and teacher outcomes? *Science Education*, *102*(2), 260–282.

Rosenberg, N. (1974). Science, invention and economic growth. *The Economic Journal*, *84*(333), 90–108. doi:10.2307/2230485

Roth, W-M. & Lee, S. (2002). Scientific literacy as collective praxis. *Public Understanding of Science, 11*, 33-56.

Rudolph, J. L. (2014). Dewey's "science as method" a century later: Reviving science education for civic ends. *American Educational Research Journal*, *51*(6), 1056–1083. doi:10.3102/0002831214554277

Ryder, J. (2001). Identifying science understanding for functional scientific literacy. *Studies in Science Education*, *36*(1), 1–44.

Schmidt, W. H., Burroughs, N. A., Zoido, P., & Houang, R. T. (2015). The role of schooling in perpetuating educational inequality. *Educational Researcher*, *44*(7), 371–386. doi:10.3102/0013189x15603982

Segarra, V. A., Hughes, N. M., Ackerman, K. M., Grider, M. H., Lyda, T., & Vigueira, P. A. (2018). Student performance on the test of scientific literacy skills (TOSLS) does not change with assignment of a low-stakes grade. *BMC Research Notes*, *11*(1), 422. doi:10.1186/s13104-018-3545-9

Servicemen's Readjustment Act of 1944, Pub. L. No. 78-346, § 78. (1944).

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

Shaw, J. M., Lyon, E. G., Stoddart, T., Mosqueda, E., & Menon, P. (2014). Improving science and literacy learning for English language learners: Evidence from a pre-service teacher preparation intervention. *Journal of Science Teacher Education*, *25*(5), 621–643. doi:10.1007/s10972-013-9376-6

Shen, B. S. P. (1975). Views: Science literacy: Public understanding of science is becoming vitally needed in developing and industrialized countries alike. *American Scientist*, *63*(3), 265–268.

Sheppard, B. H., Hartwick, J., Warshaw, P. (1988). The theory of reasoned action: A meta-analysis of past research with recommendations for modifications and future research. *Journal Consumer Research*, *15*(3), 325–343.

Shortlidge, E. E., & Brownell, S. E. (2016). How to assess your CURE: A practical guide for instructors of course-based undergraduate research experiences. *Journal of Microbiology & Biology Education*, *17*(3), 399–408. doi:10.1128/jmbe.v17i3.1103

Shrigley, R. L. (1990). Attitude and behaviour are correlates. *Journal of Research in Science Teaching*, *27*(2)97–113. doi:1002/tea.3660270203

Simpson, R. D., & Troost, K. M. (1982). Influences on commitment to and learning of science among adolescent students. *Science Education*, *66*(5), 763–781.

Sinatra, G., Kienhues, D., & Hofer, B. (2014). Addressing challenges to public understanding of science: Epistemic cognition, motivated reasoning, and conceptual change. *Educational Psychologist*, *49*(2), 123–138. doi:10.1080/00461520.2014.916216

Sireci, S. G. (1998). Gathering and analyzing content validity data. *Educational Assessment*, *5*(4), 299–321.

Solmaz, A. (2015). An analysis of the relationship between high school students' self-efficacy, metacognitive strategy use and their academic motivation to learn biology. *Journal of Education and Training Studies*, *4*(2), 53–59. doi:10.11114/jets.v4i2.1113

Spellman, J. E., & Oliver, J. S. (2001). The relationship between attitude toward science with enrollment in a 4X4 block schedule. *Proceedings of the Annual Meeting of the Association for the Education of Teachers of Science*. Costa Mesa, CA. Retrieved from https://files.eric.ed.gov/fulltext/ED472914.pdf

Stedman, L. C. (1994). The Sandia report and U.S. achievement: An assessment. *The Journal of Educational Research*, *87*(3), 133–146. doi:10.1080/00220671.1994.9941235

Summers, R., & Abd-El-Khalick, F. (2018). Development and validation of an instrument to assess student attitudes toward science across grades 5 through 10: Cross sectional instrument development. *Journal of Research in Science Teaching*, *55*(2), 172–205. doi:10.1002/tea.21416

Talton, E. L., & Simpson, R. D. (1987). Relationships of attitude toward classroom environment with attitude toward and achievement in science among tenth grade biology students. *Journal of Research in Science Teaching*, *24*(6), 507–525.

Tang, K. S. (2015). Reconceptualising science education practices from new literacies research. *Science Education International*, *26*(3), 307–324.

Texas Academic Performance Report. (2018, December). Retrieved from https://rptsvr1.tea.texas.gov/perfreport/tapr/2018/srch.html?srch=C

Texas Administrative Code. (2014). Commissioner's Rules Concerning Educator Standards §149.1001(b)(3).

Texas Education Agency. (2018, May). *STAAR biology blueprint*. Retrieved from https://tea.texas.gov/sites/default/files/Blueprint STAAR Biology 2018.pdf

Texas Education Agency. (2019a). *State of Texas assessments of academic readiness summary report: Algebra I*. Retrieved from https://tea.texas.gov/sites/default/files/Algebra%201%20Statewide_FinalTX_STAAREOC_Spr19_Summary_Reports_tagged_Part1.pdf

Texas Education Agency. (2019b). *State of Texas assessments of academic readiness summary report: Biology*. Retrieved from https://tea.texas.gov/sites/default/files/Biology%20Statewide_FinalTX_STAAREOC_Spr19_Summary_Reports_tagged_Part2.pdf

Texas Education Agency. (2019c). *State of Texas assessments of academic readiness summary report: English I*. Retrieved from https://tea.texas.gov/sites/default/files/English%20I%20Statewide_FinalTX_STAAREOC_Spr19_Summary_Reports_tagged_Part3.pdf

Texas Education Agency. (2020). *Becoming a classroom teacher in Texas*. Retrieved from https://tea.texas.gov/Texas_Educators/Certification/Initial_Certification/Becoming_a_Classroom_Teacher_in_Texas/

Texas Education Code. (1995). Title 2: Public Education Subtitle FP Curriculum, Programs and Services. §29.052(1). Retrieved from https://statutes.capitol.texas.gov/Docs/ED/htm/ED.29.htm#29.052

Texas Essential Knowledge and Skills for English Language Arts and Reading, 19 Tex. Admin. Code § 110.11 (2010). Retrieved from http://ritter.tea.state.tx.us/rules/tac/chapter110/ch110c.html

Texas Essential Knowledge and Skills for Mathematics, 19 Tex. Admin. Code §111.39 (2015). Retrieved from http://ritter.tea.state.tx.us/rules/tac/chapter111/ch111c.html

Texas Essential Knowledge and Skills for Science, 19 Tex. Admin. Code §112.34 (2018). Retrieved from http://ritter.tea.state.tx.us/rules/tac/chapter112/ch112c.html

Thompson, T. (2014). Science literacy. *Encyclopedia of health communication*. doi:10.4135/9781483346427.n478

Tong, F., Irby, B. J., Lara-Alecio, R., Guerrero, C., Fan, Y., & Huerta, M. (2014). A randomized study of a literacy-integrated science intervention for low-socio-economic status middle school students: Findings from first-year implementation. *International Journal of Science Education*, *36*(12), 2083–2109. doi:10.1080/09500693.2014.883107

Turmo, A. (2004). Scientific literacy and socio-economic background among 15-year-olds-anordic perspective. *Scandinavian Journal of Educational Research*, *48*(3), 287–305. doi:10.1080/00313830410001695745

United Nations Education Scientific and Cultural Organization. (2006). *Understandings of literacy: Education for all global monitoring report*. Retrieved from https://reliefweb.int/report/world/education-all-global-monitoring-report-2006-literacy-life

U.S. Department of Commerce. (1979). *Historical statistics of the United States: Colonial times to 1970*. Retrieved from https://www.census.gov/library/publications/1975/compendia/hist_stats_colonial-1970.html

U.S. Department of Education. (2012). *Highlights from TIMSS 2011*. Retrieved from https://nces.ed.gov/pubs2013/2013009rev.pdf

U.S. Department of Education. (2017). *Developmental education: Challenges and strategies for reform*. Retrieved from https://powerofcommunity.force.com/completionbydesign/s/article/Developmental-Education-Challenges-and-Strategies-for-Reform

U.S. Department of Education. (2018, May 14). *An overview of the U.S. Department of Education*. Retrieved from https://www2.ed.gov/about/overview/focus/what.html

U.S. Department of Education. (2019). *National assessment of educational progress data explorer.* Retrieved from https://www.nationsreportcard.gov/ndecore/xplore/ltt

Vacca, R. T., & Alvermann, D. E. (1998). The crisis in adolescent literacy: Is it real or imagined? *NASSP Bulletin*, *82*(600), 4–9. doi:10.1177/019263659808260003

Waldo, J. T. (2015). Application of the Test of Scientific Literacy Skills in the Assessment of a General Education Natural Science Program. *The Journal of General Education*, *63*(1), 1–14. doi:10.1353/jge.2014.0007

Walsh, C., Quinn, K. N., Wieman, C., & Holmes, N. G. (2019). Quantifying critical thinking: Development and validation of the physics lab inventory of critical thinking (PLIC). *Physical Review Physics Education Research*, *15*(1), 1–17.

Wang, J. (2005). Relationship between mathematics and science achievement at the 8th grade. *International Online Journal of Science and Math*, 5(1), 1–17.

Wang, J., & Ma, X. (2016). An examination of plausible score correlation from the trend in mathematics and science study. *Athens Journal of Education*, *3*(4), 301–312.

Weber, B. A., Yarandi, H., Rowe, M. A., & Weber, J. P. (2005). A comparison study: Paper-based versus web-based data collection and management. *Applied Nursing Research*, *18*(3), 182–185. doi:10.1016/j.apnr.2004.11.003

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

Weiner, B. (1990). History of motivational research in education. *Journal of Educational Psychology*, *82*(4), 616–622. doi:10.1037/0022-0663.82.4.616

Wenning, C. J. (2006). Assessing nature-of-science literacy as one component of scientific literacy. *Journal of Physics Teacher Education Online*, *3*(4), 3–14.

Wetterstrand, K. A. (Ed.). (2016, July 16). The cost of sequencing a human genome. *National human genome research institute*. Retrieved from https://www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost

Yore, L. D., & Treagust, D. F. (2006). Current realities and future possibilities: Language and science literacy—empowering research and informing instruction. *International Journal of Science Education*, *28*(2–3), 291–314. doi:10.1080/09500690500336973

APPENDICES

APPENDIX A

BEHAVIORAL, RELATED ATTITUDE, AND INTENTIONS TOWARD SCIENCE (BRAINS) INSTRUMENT

BEHAVIORAL, RELATED ATTITUDE, AND INTENTIONS TOWARD SCIENCE (BRAINS) INSTRUMENT

PUBLISHED IN THE JOURNAL OF RESEARCH FOR SCIENCE TEACHING (DOI: 10.1002/tea.21416)
DIRECT QUESTIONS ABOUT DESIGN OR USAGE TO RYAN SUMMERS (RYAN.SUMMERS@UND.EDU)

BEHAVIORS, RELATED ATTITUDES, AND INTENTIONS TOWARD SCIENCE (BRAINS)

Survey For Precollege Students in Grades 5 - 12

DEVELOPED AS PART OF THE QIAS PROJECT WITH GENEROUS CONTRIBUTIONS FROM THE QATAR FOUNDATION

PUBLISHED IN THE JOURNAL OF RESEARCH FOR SCIENCE TEACHING (DOI: 10.1002/tea.21416)
DIRECT QUESTIONS ABOUT DESIGN OR USAGE TO RYAN SUMMERS (RYAN.SUMMERS@UND.EDU)

1. Date		2. School name						
3. Grade level	☐ 5	☐ 6	☐ 7	☐ 8	☐ 9	☐ 10	☐ 11	☐ 12
5. Age		6. Gender	☐ Male	☐ Female	7. Section			

8. Which of the following best describes you? Fill in one or more ovals.	9. How often do people in your home talk to each other in a language **other** than English (e.g., Spanish or Chinese)?
☐ White	
☐ Black or African American	☐ Never
☐ Asian	☐ Once in a while
☐ Hispanic or Latino	☐ About half of the time
☐ American Indian or Alaska Native	☐ Most of the time
☐ Native Hawaiian or other Pacific Islander	☐ All the time
☐ I don't know or I prefer not to respond	

10. How far in school did your father go?	11. How far in school did your mother go?
☐ He did not finish high school	☐ She did not finish high school
☐ He finished high school	☐ She finished high school
☐ He got a vocational diploma	☐ She got a vocational diploma
☐ He graduated from a community college	☐ She graduated from a community college
☐ He graduated from a university	☐ She graduated from a university
☐ I do not know	☐ I do not know

12. How often do you talk about things you learn at school with someone in your family?				
☐ Never	☐ Once every few weeks	☐ Once a week	☐ Two or three times a week	☐ Every day

13. Is there a computer at home that you can use?	☐ Yes	☐ No		14. Can you access the internet at home?	☐ Yes	☐ No

15. In my opinion, my science grades are:	☐ Not so good	☐ Average	☐ Good	☐ Very good	☐ Excellent

PUBLISHED IN THE JOURNAL OF RESEARCH FOR SCIENCE TEACHING (DOI: 10.1002/tea.21416)
DIRECT QUESTIONS ABOUT DESIGN OR USAGE TO RYAN SUMMERS (RYAN.SUMMERS@UND.EDU)

Instructions

There are no "right" or "wrong" answers to the following questions. We are simply interested in your feelings about a number of issues related to science and science learning.

- Indicate the extent to which you agree or disagree with each the following statements.
- Place a check mark (✓) or an (✗) on the response that best represents your answer.
- Check **only one answer** for each question.

– If you "Strongly disagree" with a statement, then you should check:

Strongly disagree	Disagree	Not sure	Agree	Strongly agree
✓	○	○	○	○

– If you "Disagree" with a statement, then you should check:

Strongly disagree	Disagree	Not sure	Agree	Strongly agree
○	✓	○	○	○

– If you are "Not sure" whether you agree or disagree with a statement, then you should check:

Strongly disagree	Disagree	Not sure	Agree	Strongly agree
○	○	✓	○	○

– If you "Agree" with a statement, then you should check:

Strongly disagree	Disagree	Not sure	Agree	Strongly agree
○	○	○	✓	○

– If you "Strongly agree" with a statement, then you check:

Strongly disagree	Disagree	Not sure	Agree	Strongly agree
○	○	○	○	✓

PUBLISHED IN THE JOURNAL OF RESEARCH FOR SCIENCE TEACHING (DOI: 10.1002/tea.21416)
DIRECT QUESTIONS ABOUT DESIGN OR USAGE TO RYAN SUMMERS (RYAN.SUMMERS@UND.EDU)

	Strongly disagree	Disagree	Not sure	Agree	Strongly agree
1. I enjoy science	O	O	O	O	O
2. Scientists are highly respected	O	O	O	O	O
3. Most people should understand science because it affects their lives	O	O	O	O	O
4. I will study science if I get into college	O	O	O	O	O
5. I am sure I can do well on science tests	O	O	O	O	O
6. I usually give up when I do not understand a science concept	O	O	O	O	O
7. Science is one of the most interesting school subjects	O	O	O	O	O
8. Teachers encourage me to understand concepts in science classes	O	O	O	O	O
9. Members of my family work in scientific careers	O	O	O	O	O
10. Science is easy for me	O	O	O	O	O
11. I will not pursue a science-related career in the future	O	O	O	O	O
12. I cannot understand science even if I try hard	O	O	O	O	O
13. I will become a scientist in the future	O	O	O	O	O
14. I can understand difficult science concepts	O	O	O	O	O
15. I really enjoy science lessons	O	O	O	O	O
16. I will continue studying science after I leave school	O	O	O	O	O
17. My family encourages my interest in science	O	O	O	O	O
18. I am confident that I can understand science	O	O	O	O	O
19. We live in a better world because of science	O	O	O	O	O
20. I would enjoy working in a science-related career	O	O	O	O	O
21. Knowing science can help me make better choices about my health	O	O	O	O	O
22. My family encourages me to have a science-related career	O	O	O	O	O
23. I would like to do science experiments at home	O	O	O	O	O
24. I really like science	O	O	O	O	O
25. Scientists usually like to go to work even when they have a day off	O	O	O	O	O
26. Knowledge of science helps me protect the environment	O	O	O	O	O
27. Science will help me understand the world around me	O	O	O	O	O
28. I will take additional science courses in the future	O	O	O	O	O
29. People with science-related careers have a normal family life	O	O	O	O	O
30. I do not like science	O	O	O	O	O

APPENDIX B

TEST OF SCIENCE LITERACY SKILLS (TOSLS)

Appendix B: Test of Scientific Literacy Skills

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.

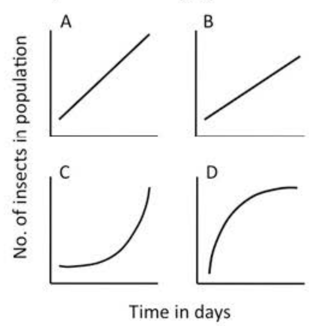
Bubble in your #ID on your scantron.

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 **<u>increase your certainty</u>** 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.

Appendix B: Test of Scientific Literacy Skills

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.

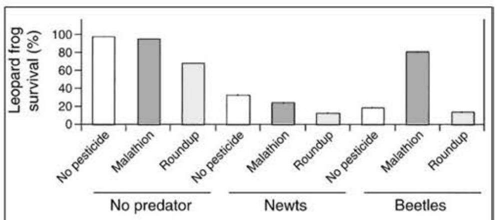


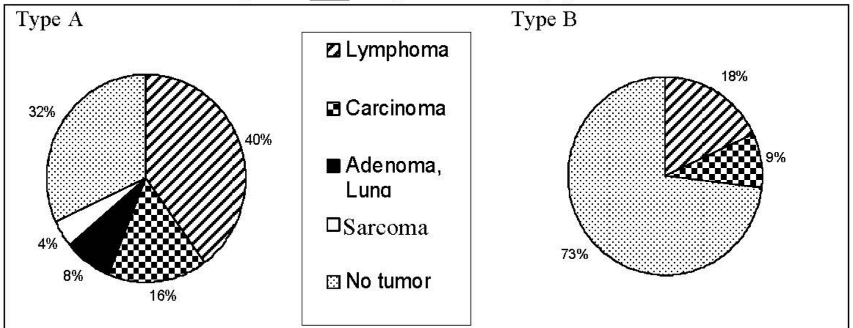
Fig. 2. Total survival of Leopard frog tadpoles living in ponds with 2 pesticides added (Malathion and Roundup) and 2 tadpole predators present (newts and beetles).

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

Appendix B: Test of Scientific Literacy Skills

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.

Appendix B: Test of Scientific Literacy Skills

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.

Appendix B: Test of Scientific Literacy Skills

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 **<u>additional evidence that supports</u>** 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.

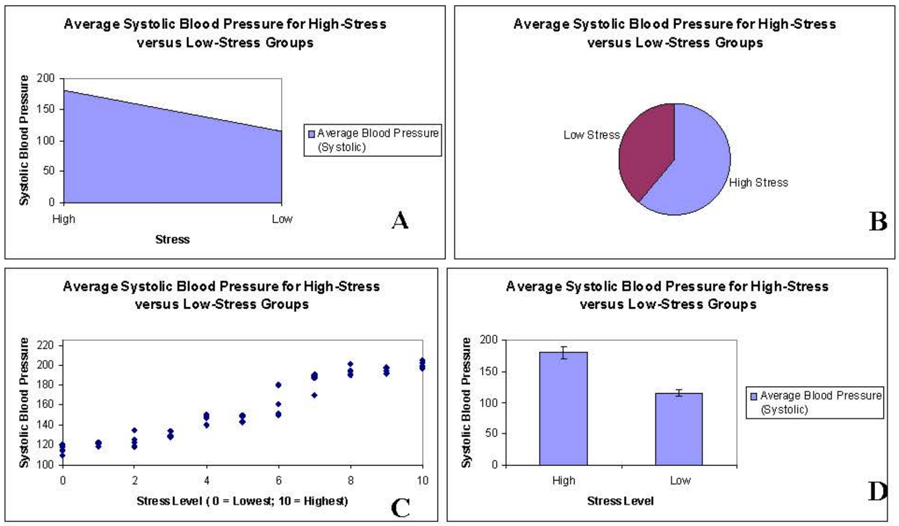
12. The excerpt above comes from what type of source of information?
    a. Primary (Research studies performed, written and then submitted for peer-review to a scientific journal.)
    b. Secondary (Reviews of several research studies written up as a summary article with references that are submitted 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

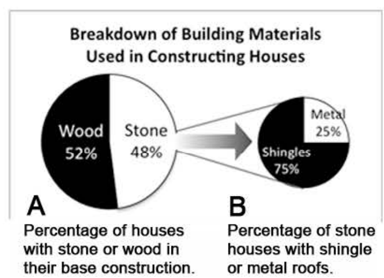
14. Which of the following attributes is **<u>not</u>** a strength of the study's research design?"
    a. Collecting data from a large sample size.
    b. Randomly sampling NYC residents.
    c. Randomly assigning participants to control and experimental groups.
    d. All of the above.

Appendix B: Test of Scientific Literacy Skills

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.

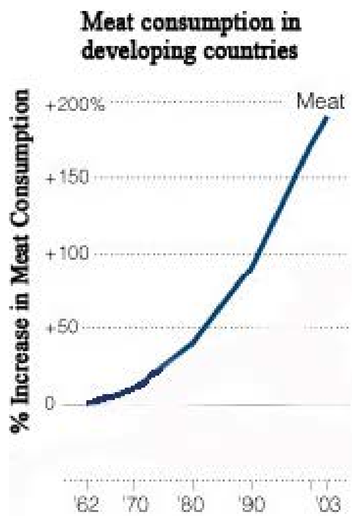


16. What proportion of houses were constructed of a stone base with a shingled roof?
   a. 25%
   b. 36%
   c. 48%
   d. Cannot be calculated without knowing the original number of survey participants.

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

Appendix B: Test of Scientific Literacy Skills

18. Which of the following is the **<u>most accurate</u>** 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 <u>equally uncertain.</u>
b. The uncertainty in the estimate of the actual mean caffeine content will be <u>smaller</u> in study 1 than in study 2.
c. The uncertainty in the estimate of the actual mean caffeine content will be <u>larger</u> 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?

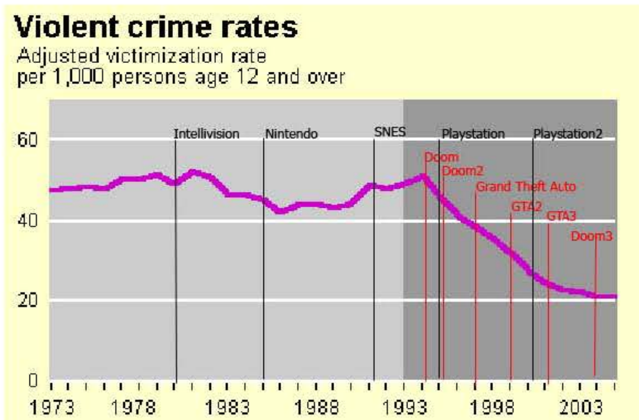
a. 40%
b. 48%
c. 60%
d. Cannot be calculated without knowing the original number of rats.

---

[1] 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

Appendix B: Test of Scientific Literacy Skills

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

Appendix B: Test of Scientific Literacy Skills

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.

Appendix B: Test of Scientific Literacy Skills

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.

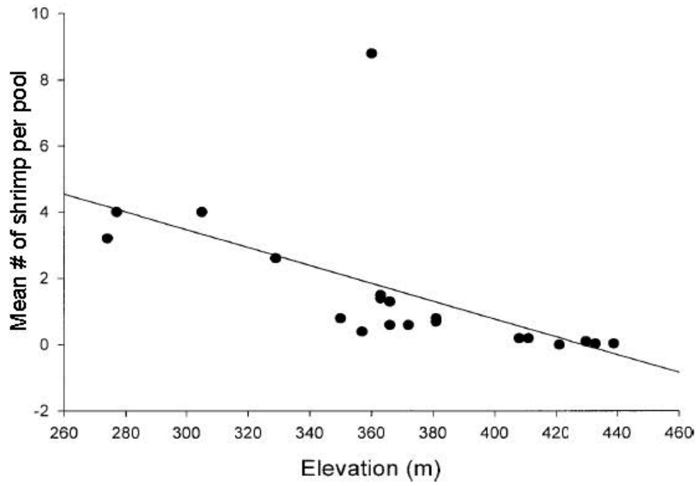


FIG. 3. Relationship between total abundance of *Macrobrachium* (1988–2002) and elevation in Quebrada Prieta.

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 C: TOSLS Skills and Answer Key (28Qs)

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 D: Student Interview Scoring Rubric

Skill	Correct reasoning	Question	Correct Answer
1 (3 Qs)	Evaluation of hypothesis with experiment; empirical test; draws conclusion that is based on strong evidence; provides reasoning for evaluating evidence; Evaluates the experimental design, recognizing confounds or need for random selection, or other independent variables	1	B
		8	D
		11	B
2 (5 Qs)	Recognize sources of bias; quoting researcher does not indicate primary source; reviews are not primary; media reports are tertiary; peer review and importance of evaluation by 3rd party experts	10	B
		12	C
		17	B
		22	C
		26	C
3 (3 Qs)	Recognize bias political or financial influences should not used to pressure findings, conclusions, reporting, or social decisions; decisions should be based on evidence; questionable ethics of publishing work that has not been verified; questionable ethics of distributing materials to bias participants; questionable ethics of rejecting studies based on controversy	5	D
		9	B
		27	B
4 (4 Qs)	No confounding factors (e.g., differences in sample size, sample selection, sample makeup); an explanation of how controls are used to mediate confounding factors; identifying strengths and weaknesses of experimental design (e.g., random assignment to control and treatment groups)	4	C
		13	D
		14	C
		25	B
5 (1Q)	Histogram is the best way to compare means	15	D
6 (4 Qs)	Given data, be able to explain what the general shape of the graph would be (exponential/logorithmic vs linear); explain why the other shapes are not correct; interpret the graph and infer cause (e.g., pesticide killed the beetles which caused more tadpoles to be eaten); extract numerical information from graph and use that to make comparisons or conclusions; interpret shape of a graph to reach a conclusion	2	C
		6	C
		7	A
		18	A
7 (3 Qs)	Solve algebraic calculations accurately	16	B
		20	B
		23	D

Appendix D: Student Interview Scoring Rubric

8 (3 Qs)	Give measure of reliability such as use of statistical tests to define probability and certainty; how sample size effects certainty; Recognize that researchers use statistics to make inferences about a population using a sample of that population	3	B
		19	C
		24	D
9 (2 Qs)	Recognize or use reasoning to explain that correlation does not imply causation; using information from a graph to explain why they chose an answer that they did (e.g., the graph showed elevation and mean number of shrimp rather than having fewer predators)	21	C
		28	B

VITA

Jeffrey R. Chandler was born in San Antonio, TX graduating from James Madison High School. He completed a Bachelor of Science in Biochemistry from Texas A&M University – College Station in 2003. He pursued a career in high school science education teaching physics and chemistry courses at various high schools. Passionate about science and science education, he earned a Master of Science in Physics from Texas A&M University – Commerce in 2011. In 2012, he was selected as a member of the Texas CSCOPE Physics Redevelopment Committee as well as the Principal Lesson Designer. This experience with curriculum resulted in being starting Simplify Science Select that provided curriculum to teachers across the state. He has written physics assessment items from Region X test banks as well as the TExES teacher certification assessment. Eager to share the scientific approach beyond the classroom, he earned a principal certification in 2017. In 2018, he moved outside the classroom to supervise the science department as Assistant Principal. He has presented at the National Advanced Placement Conference on increasing parental involvement using technology as well as numerous state and local conferences over a variety of topics including tiered formative assessment, response to intervention, and using technology in the classroom. His research interests include science literacy, science achievement, curriculum development, attitude toward science, and parental impact on student achievement.

901 Yellowjacket Ln; Rockwall, TX 75087
jeff.r.chandler@gmail.com