Education Dissertations   Education, School of

Spring May 12th, 2017

An Assessment of Integrating Authentic Research in Undergraduate Science Curricula

Daihong Chen
Seattle Pacific University

Follow this and additional works at: https://digitalcommons.spu.edu/soe_etd

Part of the Curriculum and Instruction Commons, Educational Assessment, Evaluation, and Research Commons, and the Science and Mathematics Education Commons

An Assessment of Integrating Authentic Research in Undergraduate Science Curricula

by

Daihong Chen

Dissertation

Presented to the Faculty of the

Graduate School of Education at

Seattle Pacific University

In Partial Fulfillment of the Requirements for the

Doctor of Philosophy Degree

Seattle Pacific University

May 2017

An Assessment of Integrating Authentic Research in Undergraduate Science Curricula

By Daihong Chen

A dissertation submitted in partial fulfillment

Of the requirements for the degree of

Doctor of Education

Seattle Pacific University

2017

Approved by _____

(Rick Eigenbrood, PhD, Chairperson of the Committee)

_____

(Andrew Lumpe, PhD)

_____

(Derek Wood, PhD)

Program Authorized to Offer Degree

_____School of Education_____

Date _____MAY 2017_____

_____

(Rick Eigenbrood, Ph.D., Dean, School of Education)

Signature   Daihong Chen

Date   April 9, 2017

Table of Contents

List of Tables

List of Figures

Seattle Pacific University

Abstract

An Assessment of Integrating Authentic Research in Undergraduate Science Curricula

By Daihong Chen
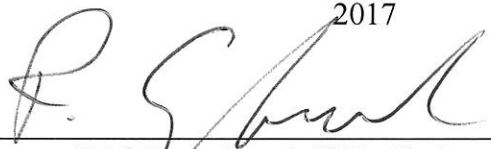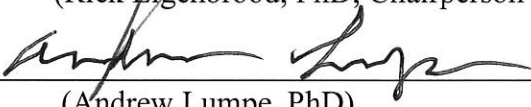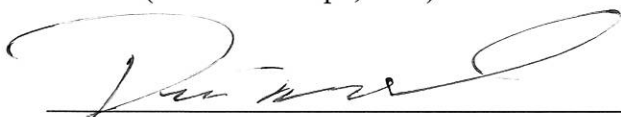
Chairperson of the Dissertation Committee:

Eigenbrood, Rick

School of Education

The purpose of this study is to investigate the impact of integrating original research projects in undergraduate science curricula on student learning outcomes. Integrating original research projects in undergraduate science curricula has been promoted as an effective approach to involving large group of students in authentic scientific inquiry. The study defines course-based undergraduate authentic research experiences or authentic scientific inquiry based on situated learning, and conducted a systematic literature review of the impact of undergraduate research experiences in science related disciplines. Based on an extensive literature review, a unique survey entitled Student Science Learning Gains (SSLG) was developed and validated to assess student self-reported science learning gains from doing authentic research integrated into undergraduate science curricula. Content validity, face validity, and construct validity were achieved via expert judge, interviews, and pilot testing. An exploratory factor analysis (principle axis factoring) with oblique rotation based on 222 responses showed that the Kaiser-Meyer-Olkin measure (KMO = . 904) verified the sampling adequacy for the analysis. The overall Cronbach's α = .94 indicated a high level of internal consistency for SSLG. The finalized SSLG consists of 29 items categorized into four constructs: self-

efficacy and attitude (8 items), concept understanding (4 items), scientific inquiry skills (14 items), and transferring (3 items), which explain 56.98% of the variance in combination. In the next step, SSLG data from 403 students who enrolled in authentic research courses were used to conduct a confirmatory factor analysis to test the six-factor model explored from the previous exploratory factor analysis. Due to high construct inter-correlations, the factorial structure of SSLG model was revised and a second order three-factor solution was tested. The second order CFA model, with three dimensions of Interest, Concept Understanding, and Inquiry Competency, had a good fit, RMSEA = .049, and CFI = .952. Scores on the scale for measuring the convergent validity, discriminant validity and the internal reliability of the higher order three-factor model yielded good estimates. After SSLG instrument was validated, relationships between authentic research experience in undergraduate courses and student scientific literacy skills were examined using path analysis. Student interest, attitudes, tool and technique skills, and communication ability were mediating variables. The latent structural equation model fit was good (RMSEA = .058, CFI = .92). The number of authentic research courses did not predict scientific literacy skills, but significantly predicted student interest (β = .16), attitudes (β =.22), tool and technique skills (β = .24), and communication skills (β = .26). Interest and communication skills had a direct relationship to scientific literacy (path coefficient = .36 and .26). Participation in authentic undergraduate research as part of a science curriculum has a moderate but positive influence on student scientific inquiry competency. The practical significance of the study, limitations, and recommendations for future research are discussed.

Chapter 1

Introduction

Statement of the Problem

The National Science Education Standards (National Research Council [NRC], 1996) stated that "scientific inquiry is at the heart of science and science learning" (p. 15). Involving whole classes of undergraduate students in research has been promoted as an effective approach to engaging students in scientific inquiry. A variety of undergraduate research projects have been launched and funded by the U.S. National Science Foundation (NSF), Research Experience for Undergraduate (REU) program, and other institutes since the 1980s. Existing studies revealed that undergraduate authentic research experiences effectively engage students in both content knowledge and procedural knowledge learning (Canaria, Schoffstall, Weiss, Henry, & Braun-Sand, 2012), and motivate students to pursue advanced education and STEM related career development (e.g., Loppatto, 2004, 2007; Urias, Gallagher, & Wartman, 2012). Nevertheless, a few issues exist in these studies in terms of the implementation and the assessment of inquiry-based instruction in undergraduate science education.

A main reason for these issues is that the conceptions and definitions of scientific inquiry have been described in a variety of ways, which results in difficulties in the understanding and the assessment of scientific inquiry teaching and learning (Hanauer, Hatfull, & Jacobs-Sera, 2009; Hofstein & Lunetta, 2003). Firstly, scientific inquiry is often misinterpreted as an instructional method that is equated with other similar teaching techniques, such as hands-on learning, learning by doing, and project-based learning that

do not guarantee meaningful inquiry is occurring (American Association for the Advancement of Science, 1993; Capps, Crawford, & Constas, 2012; NRC, 1996).

Secondly, the authenticity of the research questions that are integrated in undergraduate curriculum are not clearly clarified and sufficiently emphasized, therefore the research questions may not be investigative and meaningful, especially when the research questions are posed by students independently. Research experiences from participating in simple inquiry tasks that lack authenticity could reinforce flawed images of research practices and conceptual understanding (Chinn & Malhotra, 2002; Linn, Palmer, Baranger, Gerard, & Stone, 2015).

Thirdly, the surveys most researchers used for assessing student outcomes of undergraduate research experiences provided little information about the dimensionality and overall validity of the measurements. Developing valid and reliable instruments to specifically assess course-based undergraduate research experiences became an urgent call (Auchincloss et al., 2014). Fourthly, though benefits of undergraduate research have been reported, documented correlations do not allow a strong predictive statement to be made regarding the influence of undergraduate research on student outcomes, especially those regarding scientific literacy skills. Assessments that are founded on solid pedagogical theory and generate powerful and inferable results are rare (Auchincloss et al., 2014; Linn et al., 2015). Rigorous research that identifies ways to design meaningful research experiences and systematic assessments that document student progress with multiple indicators have been called on to address these research issues (Linn et al., 2015; Sadler & McKinney, 2010).

This study is an endeavor to bridge these research gaps in the domain of undergraduate science education. In this proposed study, integrating authentic research projects into science curriculum is suggested as an effective approach to engaging students in authentic scientific inquiry and meaningful research experiences (Chinn & Malhotra, 2002; Hume, 2009; Sadler & McKinney, 2010). Applying situated learning theory, the current study will first develop a theoretical framework that defines and clarifies authentic scientific inquiry, so as to rationalize the value and importance of integrating authentic research into science curriculum. Based on this theoretical framework, the study will then progress to investigate the impact of authentic research experience on student learning outcomes using validated instruments.

Purpose of the Study

This study is aimed to achieve two preliminary research goals: 1) clarify authentic scientific inquiry and authentic research experiences in the context of undergraduate science education; and 2) investigate the impact of integrating authentic research into undergraduate science curriculum on student learning outcomes. To address these two main research goals, this study is structured as following steps:

1. Apply the situated learning theory to define and clarify authentic scientific inquiry and authentic research experiences in undergraduate science education settings.

2. Conduct a systematic literature review of the impact of undergraduate research experiences.

3. Develop and validate an assessment instrument for assessing the impact of integrating authentic research into undergraduate science curriculum.

4. Investigate the impact of integrating authentic research into undergraduate science curriculum on student learning.

5. Discuss the scholarly significance and practical implications of the study.

Research Design and Variables

This research includes two main studies: a psychometric analysis that is used to develop and validate an instrument (Student Science Learning Gains Survey) for assessing the undergraduate science curriculum that integrates authentic research; and a predictive study that uses path analysis to investigate the predictive influence of student authentic scientific inquiry experiences on student learning outcomes. For the second research question, the level of student authentic scientific inquiry experiences, which is indicated as the number of authentic research courses a student took, is the predictive variable. The level of student scientific literacy skills is the dependent variable. Student interest, attitudes, tool and technique skills, and communication ability were mediating variables.

Method

Participants and sampling. This study will use post facto data collected from the TUE project supported by the National Science Foundation under Grant 1322848. Participants are students who enrolled in science courses that integrated authentic research projects in four higher education institutes in the United States. The TUE project used convenient sampling in data collection.

Instrumentation. The Test of Scientific Literacy Skills (TOSLS) developed and validated by Gormally, Brickman, and Lutz (2012) was used to measure student scientific literacy levels. Subscales in the Student Science Learning Gains (SSLG) survey,

developed and validated in this study, were used to measure student interest in authentic scientific practice, as well as three features of authentic scientific inquiry competency: student attitudes, tools and techniques, and communication skills (Edelson, 1998).

Data Analysis. Exploratory factor analysis (EFA) is used to understand the latent structure of variables and to identify groups of variables of the SSLG survey, so as to reduce the data set of SSLG to a more manageable size while retaining as much information as possible (Field, 2009).

After the underlying structure of the SSLG is identified, a confirmative factor analysis (CFA) is used to verify the number of underlying dimensions of the SSLG instrument that have been established on prior EFA; to identify the pattern of item-factor relationships; to find the construct validity and the reliability of SSLG; and to revise and refine the factorial structure of the SSLG (Floyd & Widaman, 1995; Hemandez, 2010).

Data collected from validated instruments will be used to conduct a path analysis (a latent structural equation model) to examine the relationships between authentic research experiences from undergraduate courses and student scientific literacy skills. Student interest, attitudes, tool and technique skills, and communication ability are mediating variables. More specifically, this study will examine the predictive influence of student authentic research experiences on student interest in science, authentic scientific inquiry competency, and student scientific literacy. In addition, this study will examine the predictive influence of student interest and scientific inquiry competency on student scientific literacy. Scientific competency in this study refers to three sub-categories: attitudes, tools and techniques, and communication skills.

Structure of the Study

       Chapter 1 provides an introduction of this study including the statement of problem, purpose of the study, and methodology. Chapter 2 discusses the definition of course-based undergraduate authentic research experiences. Chapter 3 presents a systematic literature review of the impact of undergraduate research experiences on student learning outcomes. Chapter 4 presents the context information of the undergraduate research program this study focuses on. Chapter 5 presents the process of development and validation of a new instrument for assessing student learning outcomes from participating in authentic research projects. Chapter 6 presents a path analysis that examined the predictive power of student authentic research experiences on scientific literacy skills, student interest in scientific research and scientific inquiry competency. Chapter 7 is the discussion and conclusion. The methodology and data sources used in this study are discussed in Chapter 5 and Chapter 6 respectively.

Chapter 2

Definition of Course-Based Undergraduate Authentic Research Experiences

Integrating authentic research projects in undergraduate science curriculum enables students to experience authentic scientific inquiry. Therefore, in this study, undergraduate research experiences refer to student scientific inquiry experiences. The ultimate goal of this study is to investigate the impact of integrating authentic research projects in undergraduate science curricula on student learning outcomes. As scholars pinpoint, a big obstacle to evaluating scientific inquiry-based learning is that defining scientific inquiry is problematic (Briggs, Long, & Owens, 2011). In order to conduct a systematic and effective assessment of undergraduate authentic research experiences, it is important to define course-based authentic scientific inquiry, as well as identify characteristics of authentic scientific inquiry and its' educational objectives. The purpose of this chapter is to define course-based undergraduate authentic scientific inquiry and undergraduate research experiences based on situated learning theory.

Situated Learning Theory

The primary concern of school education often seems to be the transfer of abstract, decontextualized formal concepts and knowledge (Collins, 1988). These abstract knowledge and skills are either transmitted from others, or experienced in interactions with others, through which learners internalize the knowledge. The focus on internalization interprets learning as absorbing the established knowledge as a matter of transmission and assimilation, which considers knowledge transfer a static concept and leaves the nature of the learner, the world, and the relations between them unexplored (Lave & Wenger, 1991). Learning abstract concepts independently of authentic situations

overlooks the way understanding is developed through continued, situated use. The constituent parts of all knowledge index the world and so are inextricably a product of the activity and situations in which they are continually developed (Brown, Collins, & Duguid, 1989). Interpreting learning as individual internalization of knowledge also leads to the issue that, school learning tends to occur separately from expert practice which is critical to real-world performance and is difficult to teach by lecture or explanation. When these expert knowledge and skills are taught in an abstract manner and operationalized differently from how experts and practitioners use them in daily life, it is hard for students to apply them in concrete real-world situations (Collins, 1988; Collins, Brown, & Newman, 1989; Dennen, 2004 Lave & Wenger, 1991).

Situated learning theorists assert that any type of learning is situationally grounded and manifested in collectively shared practices and identities (Lave & Wenger, 1991; Niewolny & Wilson, 2009). Situated learning is defined as "the notion of learning knowledge and skills in contexts that reflects the way the knowledge will be useful in real life" (Collins, 1988, p. 2). Within this conceptual framework, knowledge transfer is a dynamic process in which a person participates in "interactions with other people and with material and representational systems" (Greeno, 1997, p. 11), but not merely an individual cognitive process for knowledge internalization (Hotho, Saka-Helmhout, & Becker-Ritterspach, 2014). With the view that all learning activities entail social context as well as reflect social practice of human being, situated learning argues that developing learners' ability to participating in valued social practices and the identity as learners is more important than merely learning a collection of facts and procedures (Lave & Wenger, 1991; Lim, Reiser, & Olina, 2009). Through active participation in valued social

practices, learners' interest in a domain is clarified and fostered; intrinsic motivation is stimulated; as well as the meaning and purposes of learning and being a learner are configured. Meanwhile, when knowledge is learnt through continued and situated use, learners can understand the meaning of knowledge and construct individual recognition history through the interaction with the situation, which can facilitate transfer, implication, and development of the knowledge (Brown et al., 1989).

Situated learning theory involves two key components. First, situated learning theory states that, authentic situation is fundamental to all cognitive activity (Lave & Wenger, 1991). It argues that meaningful learning only takes place in authentic settings and applications, which normally involve the target knowledge and skills. Secondly, it stresses social interaction and collaboration because learning is perceived as an integral and inseparable aspect of social practice. Situated learning theory explains the nature of learning as a process of cognitive apprenticeship that occurs through legitimate peripheral participation. Learners enter in a contextual setting on the periphery as newcomers, observing the community of practice, and then gradually move toward full participation with scaffoldings provided by experienced ones. As the participation in sociocultural practices of a community increases, learners move from the role of observer to fully functioning agent, mastering the knowledge and skills, transferring from novice to expert approaching problem solving. Lave and Wenger (1991) proposed that the main functions of legitimate peripheral participation are to allow learners to understand the language and stories of a community of practice, and to learn how to communicate and negotiate both within and about the practice. Through negotiation among present and past practitioners of a community, the meanings and purposes of activities are socially constructed (Brown

et al.,1989; Niewolny & Wilson, 2009). This opportunity is rare in school learning environments due to classroom tasks mainly taking place within the culture of schools and although pedagogically useful, they fail to provide the contextual features that allow authentic activity (Brown et al.,1989).

Situated learning and cognitive apprenticeship model has been historically used in a variety of fields such as midwifery, construction, and law, for helping novices become experts through social interactions. Scholars believe that situated learning model should not be relegated to vocational and trade-based training, but be applied in K-12 and higher education (Dennen, 2004; Herrington & Oliver, 2000; Lave & Wenger, 1991). Research in situated learning has demonstrated that immersing students in authentic learning environment promotes knowledge acquisition (Lim et al., 2009; Utley, 2006; Zheng, 2010); collaboration (Shih, Chen, Wang, & Chen, 2013), and critical and metacognitive thinking skills (Herrington & Oliver, 2000). Although situated learning theory receives much interest and acclaim, the application of situated learning as an instructional model in school education remains challenging due to the lack of guidance for instructional design. For applying situated learning as a model of instruction, Herrington and Oliver (2000) developed a conceptual framework for instructional design. This practical framework consists of nine components that include authentic context, authentic activities, multiple perspectives, access to expert performances, coaching and scaffolding, opportunities for collaboration, reflection, articulation, and authentic assessment. Situated learning theory and Herrington's practical framework will be used to conceptualize and define authentic scientific inquiry experiences in undergraduate science education in the next session.

Scientific Inquiry

Traditional science instruction normally provides opportunities for students to gain established knowledge of both content and process of science through lecture and lab courses. Yet these experiences fail to allow students to solve real world science problems that are complex and ill structured, and fail to provide students with an authentic understanding and accurate perspective of scientific research (Chinn & Malhotra, 2002; Nadelson, Walters, & Waterman, 2010). Inquiry-based instruction has been stressed as an effective avenue to overcome the shortcoming of traditional science education since the 1930s (Dewey, 1933). The National Research Council (NRC), in the National Science Education Standards (NRC, 1996), stated that scientific inquiry is "at the heart of science and science learning" (p. 15) and conceptualized scientific inquiry as a series of scientific activities:

> Inquiry is a multifaceted activity that involves making observations; posing questions; examining books and other sources of information to see what is already known; planning investigations; reviewing what is already known in light of experimental evidence; using tools to gather, analyze, and interpret data; proposing answers, explanations, and predictions; and communicating results. Inquiry requires identification of assumptions, use of critical and logical thinking, and consideration of alternative explanations. (p. 23)

The definition of inquiry provided by the NRC is broadly cited as guidance in organizing scientific inquiry-based teaching and learning. Nevertheless, a few issues exist in studies related to the design, implementation and assessment of inquiry-based science learning. One major concern is that the authenticity of the inquiry is misinterpreted or

ignored when designing inquiry activities. Scientific inquiry actions are often taught as

discrete components in decontextualized laboratory settings for repetition and verification,

but key features of authentic scientific inquiry are seldom embedded in most school

inquiry tasks (Chinn & Malhotra, 2002; Hume, 2009; Wong & Hodson, 2009). Ignoring

the feature of authenticity, scientific inquiry is equated as other instructional methods

such as learning by doing, hands on experiences, problem-based learning (e.g., Bergwerff

& Warners, 2007; Nugent et al., 2012; Song & Schwenz, 2013) that do not guarantee

meaningful scientific inquiry experiences. Wong and Hodson (2009) revealed that

scientific practices described by scientists are strikingly different in contrast to the image

of science portrayed in most science curricula and textbooks. Fensham (2002) argued that

the common elements of educational scientific inquiry are not closely related with

science professions and industries. Chinn and Malhotra (2002) developed a framework

for comparing the cognitive process of authentic inquiry and simple inquiry in schools

(see Table 1). Results from an examination of science textbook using this framework

found that authentic inquiry activities are rare in school. Chinn and Malhotra (2002) are

concerned that prevalent simple school inquiry tasks may reinforce student

misunderstanding that, "science is a simple, algorithmic form of reasoning" (p. 213),

which result in a naive view of the nature of science. Corresponding to these issues,

Rudolph (2000) proposed that:

> Educators need to begin to exploit the vast literature of the science studies
>
> community, not to develop some universalist picture of science, the value of
>
> which is questionable, but to begin to understand what the various practices of
>
> science look like in all their myriad forms, in order to provide some reasonably

authentic context in which to situate the scientific knowledge claims of the

curriculum. (p. 409)

Table 1

A Comparison between Authentic Inquiry (integrating authentic research in courses) and

Simple Inquiry in School

	Authentic Inquiry	**Simple Inquiry in School**
Research questions	By researcher	Provided for students
Designing (from selecting to observing)	Purely by researchers with many variables	Ready-used design
Results	Uncertain and need inference	Certain and straightforward
Theories	Develop theories	No empirical regulation
Other reports	Relate to other reports	No need

Note. The comparison is adapted from Chinn & Malhotra (2002) p. 182-183.

Providing students with authentic scientific inquiry is therefore highlighted as a

core feature of designing and implementing inquiry-based learning. Integrating authentic

research into undergraduate science curriculum is promoted as an effective approach to

disseminating the benefits of involvement in authentic research to a larger student

population (Auchincloss et al., 2014; Wei & Woodin, 2011). Nevertheless, authentic

scientific inquiry is misinterpreted again in educational practices as the highest level of

student independence in conducting research that, "the problem procedures/design,

analysis, communication, and conclusions are for the student to design" (Buck, Bretz, &

Towns, 2008, p. 54). Integrating research experience into academic-year classes is

defined as an extension of the apprenticeship model in which students conduct

independent research projects (Wei & Woodin, 2011). When authentic scientific inquiry

is misinterpreted as the highest level of independence while students conduct research, it

leads the science education to a dangerous direction that students pose simple research questions and procedures that may not generate meaningful scientific inquiry experiences, and develop false views of the nature of science (Chinn & Malhotra, 2002).

Instead of interpreting authentic scientific inquiry as the highest level of independence that problem/question, theory/background, procedures/design, results analysis, results communication, and conclusions are not provided for students (Buck et al., 2008), the authentic learning environment and activities are highlighted as the core feature of authentic scientific inquiry in this study.

Authenticity of Scientific Inquiry in Undergraduate Science Education

Authentic scientific inquiry refers to research under study of a community of scientists currently. In authentic inquiry, research questions are formed upon elaborate theories and literatures with unknown answers, and the inquiry process requires expensive equipment, advanced techniques, and methods (Chinn & Malhotra, 2002). Scholars have claimed that effective undergraduate research experiences are resulted from engaging in authentic inquiry that makes an original intellectual or creative contribution to the discipline (Hunter, Laursen & Seymour, 2006). Nevertheless, it is challenging for undergraduate students to conduct authentic research independently in a science classroom setting due to a few realistic constrains (Chinn & Malhotra, 2002; Edelson, 1998; Lee & Songer, 2003). Firstly, schools lack the time and resources to provide such research tasks for all students to conduct independent original research in a classroom setting. Secondly, most undergraduate have not built strong theoretical knowledge and sophisticated skills and scientific reasoning to pose original research questions, to design the procedures, to analyze data, and interpret results independently.

Thirdly, undergraduate students lack attitudes of uncertainty and commitment to pursue the important scientific question independently. Students would pick up the simplest question and race through the lab with one goal in mind, to finish and leave quickly (Gormally, Brickman, Hallar, & Armstrong, 2011).

To overcome these challenges and provide students with opportunities to experience authentic inquiry, adapting and integrating original research projects into undergraduate science classrooms has been proposed as an innovative way to involve large group of students in authentic inquiry practice (Edelson, 1998; Lee & Songer, 2003; Wei & Woodin, 2011). These research projects are "designed around authentic scientific research questions, directed by a real agenda of interest to the wider scientific community, and coordinated by an active research scientist" (Hanauer et al., 2009, p. 15). Within this context, the instructor is acting as a scientist, and students are apprentices and partners in the process who will reach a deep and integral understanding of key content, reasoning skills and the core practices of science (National Science Teachers Association [NSTA], 2007). Students are considered practicing authentic scientific inquiry not because they conduct research independently, but because they are engaged in a contextualized laboratory and doing original research that is under current investigation of scientists, and the results will contribute to the application, validation, and development of scientific knowledge. Within this model, students are apprentices that start as newcomers to observe, perform tasks following guidelines and protocols, and gradually move to full participation.

The purpose of adapting and integrating original research projects into undergraduate science curricula is to create authentic learning context for students to

experience authentic inquiry. To define authentic inquiry in the context of undergraduate science education, the primary task is to build a comprehensive understanding of the authenticity of scientific inquiry in educational settings. Authenticity is a critical aspect of situated learning (Herrington & Oliver, 2000). Core to cognitive apprenticeship as a method of learning is the belief that engagement in authentic setting foster relevant, transferable learning. Different from other learning methods such as hands on learning, learning by doing, or problem-based learning, situated learning requires a deeper embedding within an authentic context. When applying situated learning theory in science education settings, it is important to clarify the notion of "authenticity". Strobel, Wang, Weber, and Dyehouse (2013) described scientific inquiry-based learnig as a form of authentic learning that focus on engaging students in expert-like activity and providing real world problems.

Scientific inquiry-based learning comprises context authenticity, tasks authenticity, and impact authenticity (Strobel et al, 2013; Dennen, 2004). Context authenticity refers to students being involved in everyday cognition that entails authentic and collaborative enviroment in which knowledge is applied in practice (Choi & Hannafin, 1995). Task authenticity means students conduct ordinary practice of the culture of scientific community (Brown et al., 1989). Impact authenticity means products of students' investigation can contribute to the community of scientists (Barab, Squire, & Dueber, 2000). In undergraduate science education settings, students experience scientific inquiry by participating in research projects that are either provided by instructors or created by students. To ensure these three dimensions of authenticity, research projects that students participate in are particularly important. Research opportunities that are

integrated in science courses should allow students to address an original research question or problem that is of interest to broader community with an outcome that is unknown both to the students and the community of scientists (Auchincloss et al., 2014).

It is undoubted that undergraduate students, or even secondary school students, are able to investigate original research questions independently and contribute to the scientific knowledge development, but it is not the case discussed in this study. The focus of this study is course-based undergraduate research experiences, which means all students enrolled in an undergraduate science course have the opportunity to practice authentic inquiry. Given limited time and resources, the insufficiency of undergraduate students' knowledge and skills, and logistical restraints, it is unrealistic that the majority of undergraduates are able to form an original research question that is investigative and valuable to the community of scientists.

Previous research warns that projects created by students are concerned with little meaning to the real world and students have to re-learn the skills when they deal with real world issues (Herrington & Oliver, 2000; Zheng, 2010). In order to engage the whole class of undergraduates in meaningful and authentic scientific inquiry, integrating research projects that are "designed around authentic scientific research questions, directed by a real agenda of interest to the wider scientific community and coordinated by an active research scientist" (Hanauer et al., 2009, p. 15) into science curriculum is a more effective, economic, and practical approach.

Bringing original research projects that instructors are currently conducting to the classroom creates an authentic and collaborative learning environment attaining context authenticity, activity authenticity, and impact authenticity (Herrington & Oliver, 2000;

Strobel et al., 2013). When original research projects are brought into undergraduate science classes, students are involved in the use of scientific practice including scientific procedures and a variety of techniques; generating new knowledge or new understanding of the world; broad relevance and importance to the community of scientists; and in collaborative and interactive work (Auchincloss et al., 2014).

The Role of Students in Authentic Inquiry-based Learning

Within the theoretical framework of situated learning and cognitive apprenticeship, another key component of effectively teaching authentic scientific inquiry in the context of undergraduate science education is to understand the role of students in the inquiry process. A critical aspect of situated learning is the notion that learning occurs through legitimate peripheral participation. Learners are treated as apprentices who observe the community of practice, assist experts with some basic tasks, and gradually become fully functional agent when the involvement in the culture increases. According to the notion of legitimate peripheral participation, students participating in authentic research projects are research apprentices who begin as observer, and then complete small tasks. As students gain experience, they are offered larger and more central tasks to complete. The authentic inquiry experiences are about both of the holistic scientific inquiry process from observing and collecting data to academic writing and presentation, and evaluating performance through the completion of small tasks (Dennen, 2004).

The focus of a cognitive apprenticeship is on developing cognitive skills through participating in authentic learning experiences. As Dennen (2004) stated, apprenticeship as a method of teaching and learning is essentially one form of social constructivist methods, which requires scaffolding, modeling, mentoring, and coaching as the means

for facilitating learning. When students enroll in a science course that integrates original

research project but are with little authentic scientific inquiry experience, especially

freshman and sophomore students, it requires instructors to provide scaffolding and

coaching at critical times (Zheng, 2010). Enkenberg (2001) proposed an instructional

strategy for guiding the teaching and learning through cognitive apprenticeship.

1. Modeling: meaning the demonstration of the temporal process of thinking.

2. Explanation: explaining why activities take place as they do.

3. Coaching: meaning the monitoring of students' activities and assisting and
   supporting them where necessary.

4. Scaffolding: meaning support of students so that they can cope with the task
   situation. The strategy also entails the gradual withdrawal of teacher from the
   process, when the students can manage on their own.

5. Reflection: the student assesses and analyses his performance.

6. Articulation: the results of reflection are put into verbal form.

7. Explorations: the students are encouraged to form hypotheses, to test them,
   and to find new ideas and viewpoints. (Enkenberg, 2001, p. 503)

Based on Enkenberg's strategy of learning and teaching through cognitive

apprenticeship, students are engaged in acts of observation, practice, and reflection.

When students increasingly gain experiences, the modeling and coaching from instructors,

as the experts, are fading gradually (Collins, 1988). Therefore, based on cognitive

apprenticeship model, the definition of authentic inquiry as the highest level of

independence of students through scientific inquiry is abandoned in this study. The role

of students in authentic inquiry, especially in the context of course-based learning

environment, is not considered as a researcher who poses their research questions and design the procedure independently, but is a research apprentice who observes, complete small tasks, and then is offered larger tasks. When students gain sufficient knowledge and skills, they are involved in more central and fuller participation, and have increased independence to design and develop an authentic research activity.

Definition of Course-based Undergraduate Authentic Inquiry

Grounded on situated theory and related research, in the study presented in this dissertation, authentic scientific inquiry in the context of undergraduate science classroom is defined as a form of original research project-based authentic learning. The aim of adapting and integrating original research projects into undergraduate science curricula is to provide whole-class students with opportunities to experience authentic inquiry, through which students are invoved in the culture of scientific community increasingly, develop the identification as scientists, and transfer from a newcomers to a full functional agents through peripheral participation. This definition interprets the authenticity of scientific inquiry as an authentic learning environment with features of context authenticity, tasks authenticity, and impact authenticity.

Context authenticity refers to bringing original research questions that are in interest of the scientists community to undergraduate classroom. Task authenticity refers to that students are modeling what professional scientists practice daily. Impact authenticity refers to that the inquiry makes an original intellectual or creative contribution to the discipline. The role of students in this definition is interpreted as apprentices participating in the practice of the culture of scientific community

progressively with the modeling, mentoring, coaching and scaffolding provided by experts, rather than the degree of independence of students in the inquiry process.

The original project-based learning design in nature meets the features of instructional design framework developed by Herrington and Oliver (2000) for effectively implementing situated learning in school. First, original research project-based learning provides authentic context that reflects the way knowledge is used in real life. Secondly, students practice authentic scientific inquiry process and skills when they are engaged in original research projects. Thirdly, the original research projects are currently under conduction by scientists who are also instructors to the courses. It allows easy access to expert performance and the modelling of processes. Fourthly, original research aims to generate new scientific knowledge and application, which provides opportunities to experience multiple roles and perspectives in problem solving. Fifthly, original research requires collaboration in nature. Sixthly, it promotes reflection with providing students with the opportunities to compare their performances and results with experts and peers. Seventhly, students interpret and negotiate their findings via academic writing or presentation, which promote articulation to enable tacit knowledge to be made explicit. The other two components of Herrington and Oliver's (2000) framework are providing coaching and scaffolding at critical times, and integrated assessments within the tasks, which requires pedagogical strategies and efforts from individual instructor.

Chapter 3

Systematic Literature Review of the Assessment of Undergraduate Research Experiences

Integrating research into undergraduate science courses has been promoted as an effective way to teach scientific inquiry and enhance science education. The purpose of this study is to develop an instrument and to assess the impact of integrating authentic research into undergraduate science courses, so that the review centers on literatures regarding the assessment of undergraduate research experiences in science related disciplines.

This study applies systematic literature review, which is a means of identifying, evaluating, and interpreting all available research relevant to a particular research question. Systematic review methodology is distinguished from narrative reviews of the literature in two aspects. First, systematic review emphasizes transparent, structured, and comprehensive approaches to searching the literature. Second, it requires for formal synthesis of research findings. Nevertheless, there appears relatively little use of the systematic review methodology within the higher education sector (Bearman et al., 2012). The purposes for undertaking a systematic literature review in this study are to summarize the existing evidence concerning the benefits and limitations of integrating research projects in undergraduate science education; to identify gaps in current research in order to suggest areas for further investigation; and to provide a framework /background in order to appropriately position new research activities.

Systematic Literature Review Design and Process

The systematic literature review of the impact of undergraduate research experiences follows the guidance provided by Petticrew and Roberts (2006). According to Petticrew and Roberts, conducting a systematic literature review includes seven stages in conducting a systematic review:

1. Clearly define research questions the systematic literature review is expected to answer.

2. Determine the types of studies for answering systematic literature review research questions.

3. Conduct a comprehensive literature search to locate studies.

4. Screen the results of search according to inclusion criteria and exclusion criteria.

5. Critically appraise the included studies.

6. Synthesize the studies and assess heterogeneity among the study findings.

7. Conclusions and recommendations.

Research questions. Formulating research questions is the most important part in systematic literature review. The research questions are not necessarily the same as research questions addressed in the current study, but are used to guide the literature search process and the extraction process. Data analysis aims to answer systematic literature review research questions. Petticrew and Roberts (2006) suggested that the research questions about effectiveness of a treatment should be formulated according to five elements known as PICOC. The first element is population, which refers to the target group for the intervention. The second element is intervention, which refers to what

intervention this study is interested in reviewing. The third one is comparison, which

refers to with which the intervention is being compared to. The forth one is outcome,

which refers to the effect of the intervention. The last one is context, which refers to

within which the intervention is delivered. The five elements of the systematic literature

review are presented in Table 2.

Table 2

Five Elements of Undergraduate Authentic Research Experiences

Element	Description
Population	Undergraduate students who have research experiences in science related domains.
Intervention	Undergraduate research experiences.
Comparison	Lecture-based and standard lab instruction.
Outcomes	Impact of undergraduate research experiences on student learning outcomes.
Context	All empirical studies about undergraduate research experiences or scientific inquiry experiences.

Guided with the PICOC, research questions of the systematic literature review are

formed as following:

Primary research question: What evidence is there of the impact of the research

experiences on undergraduate students' science learning outcomes, especially comparing

to the traditional science education model that is lecture-based and uses standard lab?

Sub-research questions:

1. How were the undergraduate research programs (e.g., course based model,
    internship model, mentored model, summer research program) implemented?

2. What is the authenticity of undergraduate research projects in URE studies?

3. How were the students' learning gains measured in URE studies?

4. What instruments were used in assessing students' learning gains from undergraduate research experiences? What instruments are validated?

Literature search. This literature search was limited to English-language abstracts of articles published between January 1950 and April 2016 using the key words of "undergraduate research experiences", "undergraduate scientific inquiry", "authentic research", "scientific inquiry", "authentic scientific inquiry". For refining search results, the key word "science" was added to "undergraduate research experiences". Electronic databases searched are presented in Table 3. The results of each search string were assessed on screen to ascertain that studies were likely to meet inclusion and exclusion criteria that were derived from concepts inherent in both of the primary review questions and sub-questions. The inclusion and exclusion criteria are presented in the Table 4.

Table 3

Electronic Databases Searched

Database Name
ERIC (U.S. Dept. of Education)
MEDLINE/PubMed (NLM)
Directory of Open Access Journals (DOAJ)
SpringerLink
JSTOR Archival Journals
Dialnet
PMC (PubMed Central)
DTIC Technical Reports (U.S. Defense Technical Information Center)
EScholarship
SciELO Brazil (Scientific Electronic Library Online)
BioMed Central
SwePub (National Library of Sweden)
DiVA - Academic Archive Online
SpringerLink Open Access
ArXiv
SwePub (National Library of Sweden)- Free access
HathiTrust Digital Library
UNT Digital Library
Medknow Publications
UBIRA eTheses

Table 4

The Inclusion and Exclusion Criteria

Inclusion criteria	Exclusion criteria
- Written in English - Empirical research - Peer-reviewed papers published between 1950-2016 - Focus on undergraduate research programs or scientific inquiry-based learning - Studies are about undergraduate level science related disciplines - Include assessment of student learning outcomes	- Not written in English - Not based on empirical research - Based on single person opinion - Not focus on undergraduate research or scientific inquiry – based learning - Not related to science or science related fields - Provide little information about the assessment of student learning outcomes - Books, dissertations and book reviews were excluded, due to time and resource limitations

Data extraction and quality of study assessment. A framework was designed for extracting, assessing and analyzing the data contained within the included studies (an example is presented in the Appendix A). The framework was used to support the process of synthesizing and reporting the review findings and report writing, and also used to reduce any bias from the processes that mediate the research process and production. The data extraction framework comprises following sessions: bibliographic information; purpose of the study, the research project described in the study, instructional design (how research projects are delivered), authenticity of the research project, assessment methodology, sample size, validation of the measures, and the impact on student learning,

and the limitation. The framework is designed to ensure that the data is extracted

consistently. The quality of the studies and the weight of evidence within each study were

assessed by an analysis of the strength and limitations of the empirical studies. Three

components were identified to assess the quality of the studies: the soundness of the

studies, the appropriateness of the research design and analysis, and the relevance of the

study topic focus (e.g., sample, measure, instructional settings, and authenticity) to the

review questions. Judgement of overall weight of evidence (WoE) based on the

assessments according to criterion created by Davies et al. (2013) (see Table 5).

Table 5

Criteria for Judging "Weight of Evidence"

Level/Criterion	Methodological quality	Methodological relevance	Topic relevance
1: Excellent	Excellent research design justifying all decisions taken (e.g. sample, instruments, analysis. Clear evidence of measures taken to maximize validity and reliability).	Research questions clearly stated. Methodology is highly relevant to RQs and answers them in detail.	Study is very closely aligned to one of the key review questions and provides very strong evidence upon which to base future policy/action.
2: Good	Research design clearly stated with evidence of sensible decisions taken to provide valid and reliable findings.	Research questions are explicit or can be deduced from text. Findings address RQs.	Study is broadly in line with one of the key review questions and provides useful evidence.
3: Satisfactory	Research design may be implicit but appears sensible and likely to yield useful data.	RQs implicit but appear to be broadly matched by research design and findings.	At least part of the study findings is relevant to one of the key review questions.
4: Inadequate	Research design not stated and contains flaws.	RQs not stated or not matched by design.	Study does not address key questions.

Search Results

  The search process and results are presented in Figure 1. Sixty-seven primary studies appeared to meet the inclusion criteria after fully paper screening. Twenty-three papers were excluded after the first fully paper screening because these studies merely discussed undergraduate research and assessment theoretically, or introduced an undergraduate research program but did not include any information about the assessment of student learning.

Initial search ($n = 2,404$)

Further key words applied to refine search
($n = 763$)

Abstract and title screened ($n = 165$)

First round full-text screened for potential inclusion
($n = 97$)

Second round full-text screened for final inclusion
($n = 74$, assessment reviews: $n = 6$, primary studies: $n = 67$)
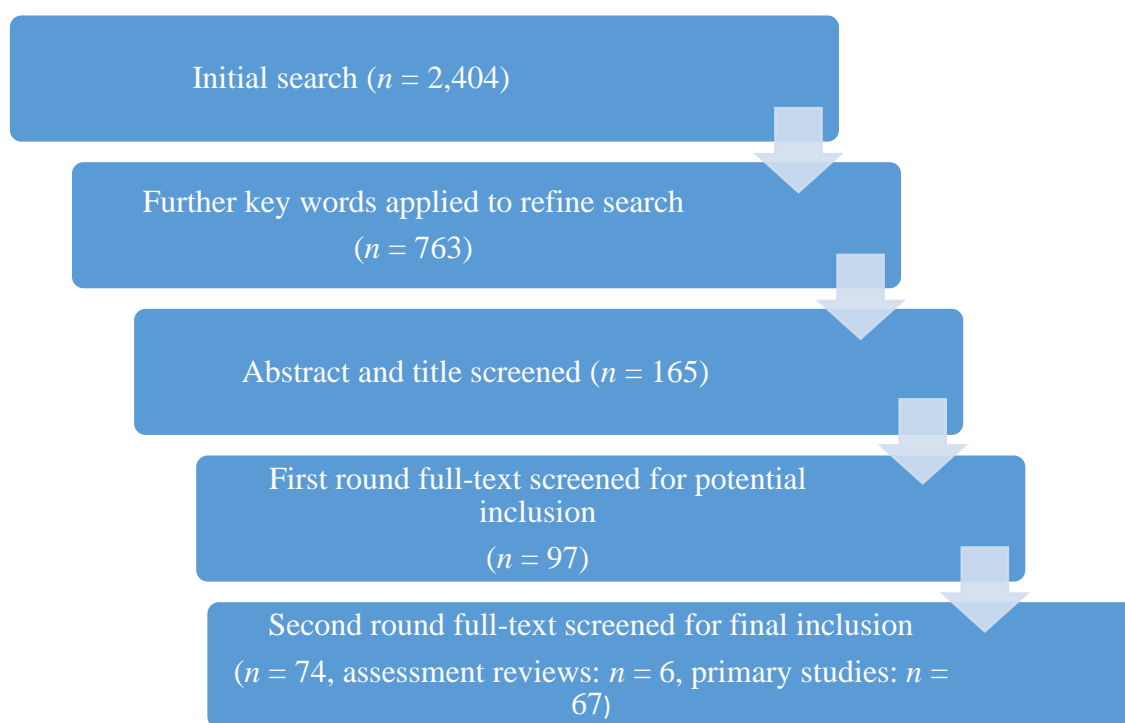
Figure 1. Search Process and Results

Review of Reviews

  Six studies (Corwin, Graham, & Dolan, 2015; Crowe & Brakke, 2008; Linn et al., 2015; Sadler, Burgin, McKinney, & Ponjuan, 2010; Sadler & McKinney, 2010; Wei & Woodin, 2011) reviewed the literature of student learning outcomes of undergraduate research experiences.

These six studies all examined the benefits of research experiences. Corwin et al. (2015) reviewed the studies of CUREs and research internships to generate a comprehensive set of outcomes of research experiences, determining the level of evidence supporting each outcome. Sadler and McKinney (2010) reviewed 20 empirical studies published between 1992 and 2007. The results of their review indicated that undergraduates tend to demonstrate learning outcomes including career aspirations, confidence, nature of science, intellectual development, content knowledge, and skills, but the extent to which these gains match expected and possible gains varies across outcomes. Sadler et al. (2010) reviewed 53 studies of scientific research apprenticeship experiences for secondary students, undergraduates and teachers, both pre-service and in-service. The review explored various learning outcomes associated with participation in research apprenticeships. These outcomes included effects of apprenticeship experiences on participant career aspirations, ideas about the nature of science (NOS), understandings of scientific content, confidence for doing science and intellectual development. Findings related to some themes (e.g., NOS understandings) supported conflicting conclusions.

In the review conducted by Crowe and Brakke (2008), the authors briefly summarized 24 studies, and stated that the assessment of undergraduate-research experience is in the early stages and encouraged more attention to assessment of outcomes. Linn et al. (2015) reviewed 60 articles published in the last five years. The authors first compared independent undergraduate research experiences and course-based undergraduate research experiences, which vary in selectivity, duration, setting, mentoring, and cost. This review synthesized the benefits of undergraduate research which include promoting persistence and identity, improving research practices,

expanding conceptual understanding, communicating the nature of science. Wei and Woodin (2011) reviewed 14 course-based research programs, and pointed out a few of the promising emerging efforts to integrate research experiences into academic-year classes.

Reviewers suggested the need for comparison study between research-based science learning and traditional lab (Linn et al., 2015). Some reviewers pointed out that the reliability of measures that identify instrument features are missing in these studies reviewed. Some of the most important variables of interest in analyses of apprenticeship programs such as nature of science (NOS) understandings, scientific content knowledge, and intellectual development are from self-reported data with little information of validity (Sadler et al., 2010).

Some researchers claimed that the reviews explored authentic research experiences as contexts for learning, and defined authentic research as "opportunities for learners to work on scientific research with practicing scientists" (Sadler & McKinney, 2010, p. 44). Nevertheless, there is little information of the examination of the authenticity of the research experiences in the reviewed studies.

Review of Primary Studies

Sixty-seven primary studies are included in this systematic literature review. Following session summarizes the content findings from accepted primary studies.

Characteristics of primary studies. Although this systematic literature review examined studies published since 1950, no studies that met the inclusion criteria were published before 2002. The publications' trend of the included studies (Figure 2) clearly shows that the number of works on assessment of undergraduate research and scientific

inquiry experiences increased remarkably since the beginning of the 1920s. Figure 3 shows the delivery model of undergraduate research experiences. Among 68 studies, 28 studies (41%) investigated the impact of undergraduate course-based research experiences. Seven (10 %) assessment studies used large-scale data crossing institutions and disciplines. Nineteen (27%) studies assessed the undergraduate student research experiences from other models (e.g., internship model, mentored model, selected student model, and summer research programs, extracurricular certification program). Fifteen studies (22%) did not provide enough information to identify the specific delivery model of undergraduate research experiences. Figure 4 displays study design of assessing the impact of undergraduate research experiences including quantitative methods, qualitative methods, and mixed methods. Nine studies (14%) did not conduct a formal assessment but only provide instructors' opinion-based commentary. One study (Lopatto, 2011) did not provide any information of data resources and analysis.
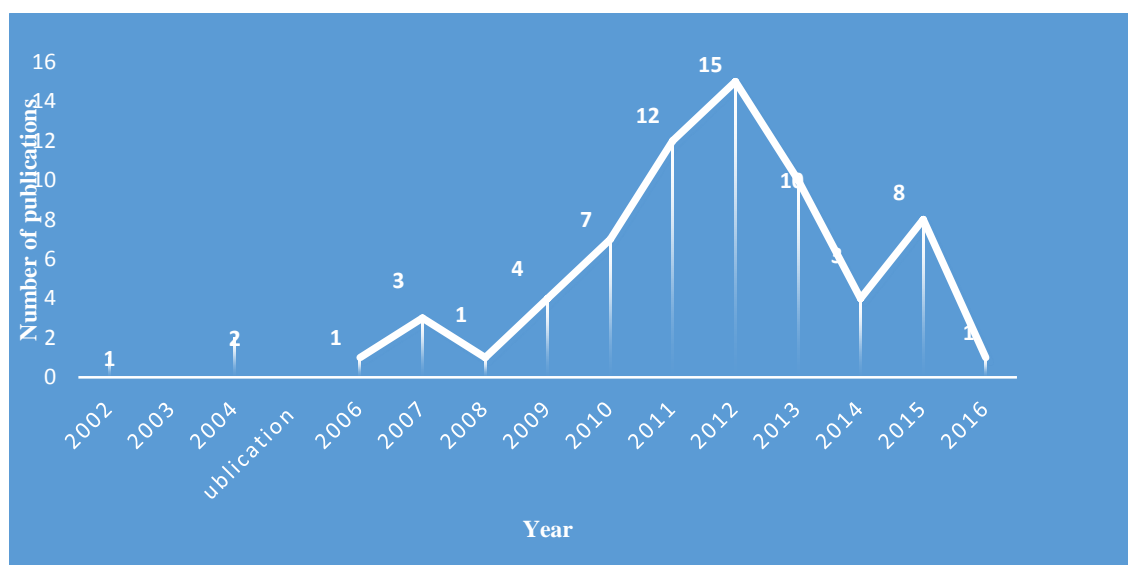
Figure 2. Trend of Publications
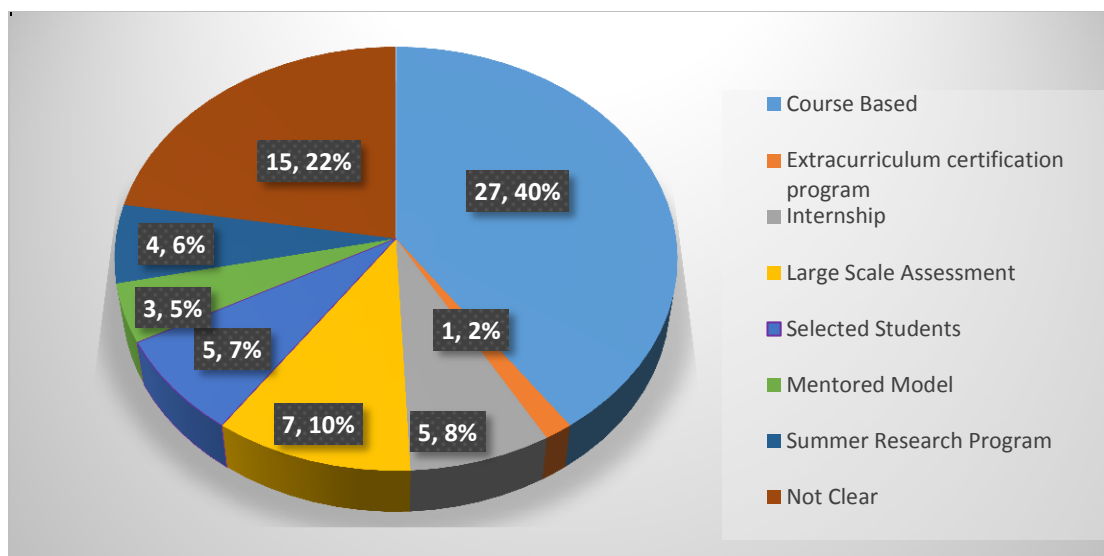
Figure 3. Undergraduate Research Experiences Delivery Approaches
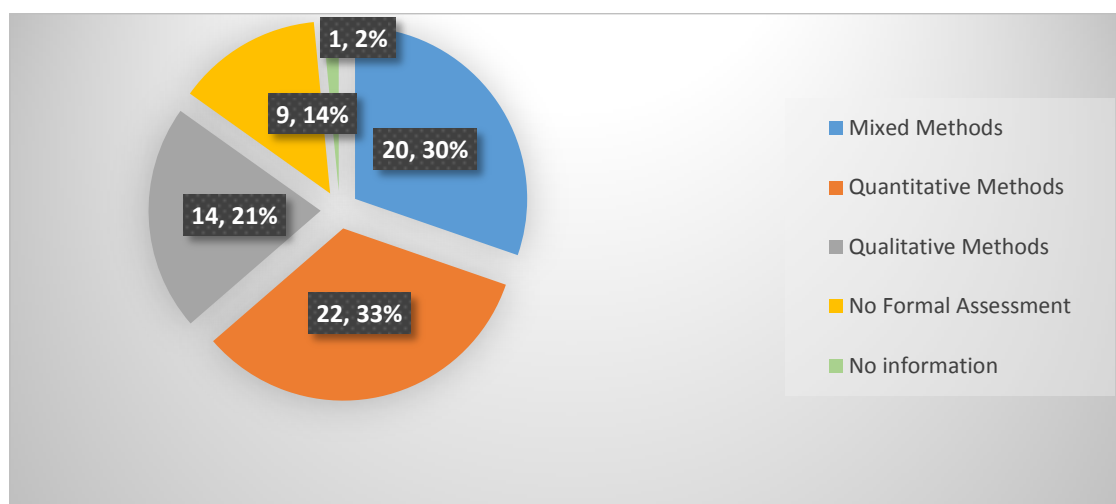
Figure 4. Assessment Study Design

Quality of the study. Figure 5 displays the quality of included studies using the criteria for judging "weight of evidence" (Davies et al., 2013, p. 83). The results showed that 13% of studies are excellent ($n = 9$), which used random controlled study design and validated instruments, or rigorous qualitative methods that ensure the validity and trustworthiness. Thirty-two percent are good ($n = 22$), which used quasi-experiment study design, comparison or correlation studies using inferential statistics, or well-designed

qualitative study. Thirty-four percent are satisfactory, which used pre-post single group

design, descriptive statistics, or narrative description of qualitative data ($n = 23$). Twenty-

one percent of the studies did not formally assess student learning outcomes, but used

instructors' opinion based commentary on students' performance and schoolwork

products ($n = 14$). Regardless of the assigned quality score, all studies are included in this

review.
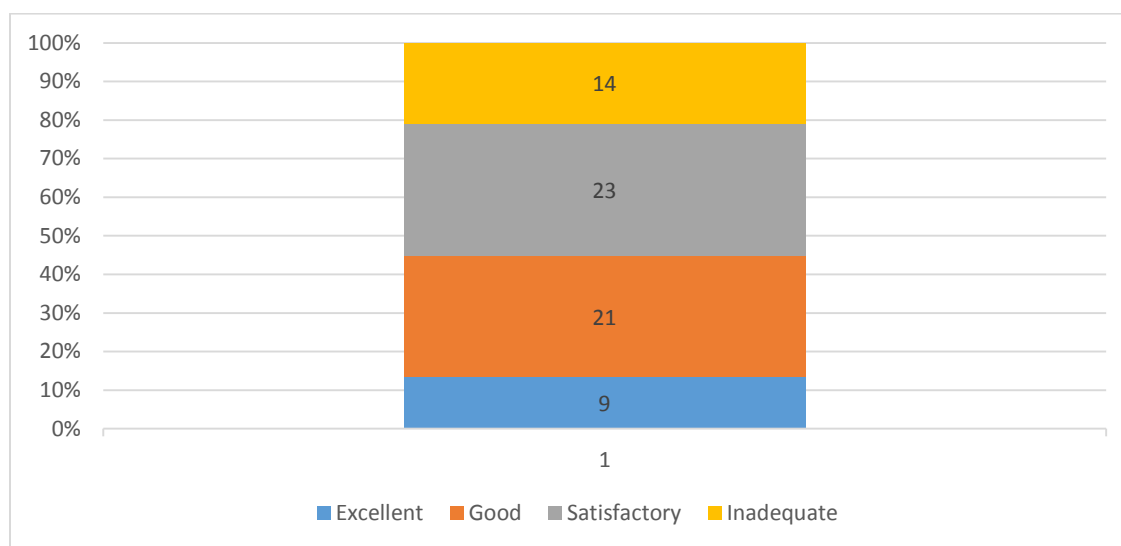
Figure 5. The Quality of Included Studies

Among studies that analyzed quantitative data, 20 studies used descriptive

statistics, and 25 studies used inferential statistics. In total, there are five control studies,

two randomly controlled study (Miller, McNeal, & Herbert, 2010; Schussler, Bautista,

Link-Pérez, Solomon & Steinly, 2013), three quasi-experiment study (Nugent et al., 2012;

Nugent, Kunz, Levy, Harwood & Carlson, 2008; Russell et al., 2015). There are six

comparison studies that investigate the differences between groups of individuals that

were not matched (Hanauer, Frederick, Fotinakes, & Strobel, 2012; Hartmann, Widner,

& Carrick, 2013; Kardash & Edwards, 2012; Luckie et al., 2012; Nadelson et al., 2010;

Thiry, Weston, Laursen, & Hunter, 2012). Seven are correlational studies (Gilmore,

Vieyra, Timmerman, Feldon, & Maher, 2015; Ing, Fung, & Kisailus, 2013; Jaarsma et al., 2009; Jones, Barlow, & Villarejo, 2010; Pender, Marcotte, Domingo & Maton, 2011; Shields, Hewitt, & North; 2010; Taraban & Logue, 2012). Seven studies used pre-post single group design (Brongo & Norman, 2011; Campbell, Wolf, Der, Packenham, & Abd-Hamid, 2012; Davidson & Palermo, 2015; Lustick, 2009; Naug, Colson & Donner, 2012; Woodzicka, Ford, Caudill, & Ohanmamooreni, 2015; Zimbardi, Bugarcic, Colthorpe, Good, & Lluka, 2013).

Eleven studies used validated instruments (provided information about the reliability or validity of the instruments) or validated rubrics. One study used one validated measure but there was little information about the validity of the other measure used in this study. Four studies used university student records including SAT, GPA, application ratings, transcripts, and standardized test scores. Twelve studies used validated instruments or university student records and inferential statistics. Among these, only four studies investigated course-based research experiences. The rest of studies that used surveys for data collection used researcher designed surveys or existing surveys but provided little information about the validity and reliability of the instruments.

The examination of the quality of included studies indicates that high quality study that used solid methods and validated instruments are scarce, particularly in the field of assessment of course-based undergraduate research experiences.

Authenticity of undergraduate research experiences. Authentic situation is fundamental for any cognitive activity (Lave & Wenger, 1991). In this study, the term of authentic scientific inquiry or authentic research experiences is defined as a form of original research project-based authentic learning with the features of context authenticity,

activity authenticity, and impact authenticity, which involves students in the culture of

scientific community increasingly and transfers students from a newcomers to a full

functional agents through peripheral participation. Participating in original research

projects, students address a research question that is of interest to the broader community

with an unknown outcome both to the students and instructors, and the research results

would contribute to the validation and development of scientific knowledge (Auchincloss

et al., 2014). Based on this definition, the authenticity of research projects described in

included studies is examined using the criteria described in Table 6). These criteria are

adopted and modified from the rubrics that were validated by Strobel et al. (2013).

Table 6

Criteria for Rating the Authenticity of Undergraduate Research Experiences

Type of Authenticity	Rating Criteria
Context authenticity	Real-world context / future professional situation Research question is of interest to the broader community of scientists Complete task-environments Ill-structured, non contrived problems with ambiguous data
Task authenticity	Observing and practicing what scientists do when they conduct research Suspension of disbelief Interaction among learners and senior researchers.
Impact authenticity	Making original intellectual or creative contribution to the discipline. Values definsible in objective terms Classroom-professional community balance Results dissemination in professional conferences or journal publications

Note. The criteria are adopted from Strobel et al., 2013, p. 148.

Research projects described in 23 studies are rated as context authenticity, task authenticity, and impact authenticity (Adedokun & Burgess, 2011; Barker, 2009; Chung & Behan, 2010; Coverdale, 2002; Culp & Urtel, 2013; Dillner, Ferrante, Fitzgerald, & Schroeder, 2011; Hanauer et al., 2012; Hunter, Laursen & Seymour, 2006; Ing et al., 2013; Jaarsma et al., 2009; Jansen et al., 2015; Miller, Hamel, Holmes, Helmey-Hartman, & Lopatto, 2013; Pacifici & Thomson; 2011; Quardokus, Lasher-Trapp, & Riggs, 2012; Russell et al., 2015; Canaria, Schoffstall, Weiss, Henry, & Braun-Sand, 2012; Shanle, Tsun, & Strahl, 2016; Thiry et al., 2012; Urias et al, 2012; Wagner et al., 2010; Willis, Krueger, & Kendrick, 2013; Zhan, 2014). Research projects described in three studies showed context authenticity, task authenticity, but not impact authenticity (Ellis-Monaghan & Pangborn; 2013; Iimoto & Frederick, 2011; Miller et al., 2010; Woodzicha et al., 2015). Examples of authentic research projects and non authentic research projects are presented in Table 7.

Table 7

Examples of Authentic Research Project and Non-Authentic Research Project

Types of Authenticity	Examples of Research
Context authenticity, task authenticity, impact authenticity	1. Each module is separated into three sections consisting of 1) introductory material, including background information and skills the students need to perform the research project; 2) the research methods, data acquisition and analysis, and new results that are useable by the research module author; and 3) an open-inquiry project, where students develop and execute their own research idea that extends the original project. Students complete a portion of a publishable research project and contribute to future work of the module author (summer research program).

	(Quardokus et al., 2012)
	2. Faculty member is responsible for obtaining preliminary funding, the initial research question has been developed before bringing the undergraduate students into work on the project. (Culp & Urtel, 2013; Shanle et al., 2016)
Context authenticity, task authenticity, no impact authenticity	Experimental group activities in this study brought real-world issues and exposure to ill-constrained problems common to coastal systems into the classroom through manipulation of large-scale data-sets and the use of multiple representations. (Miller et al., 2010).
Non-authentic research projects	Students choose the research question, variables, and protocol and explain their results in light of other studies and theories (Gormally et al., 2011). Learning what is authentic research and inquiry through presentation given by researchers and interviewing researchers (Behar-Horenstein & Johnson, 2010).

The examination of the authenticity of the research projects found that studies and assessment of course-based undergraduate authentic research programs are scarce, especially using validated instruments and sophisticated research methods. In seven studies, course based research porojects are original research with context authenticity, task authenticity. In three studies, course based research projects are rated with context authenticity, task authenticity, but not impact authenticity. Among these 10 studies in terms of course-based research experiences, one was randomly controlled study that quantified qualitative data using validated rubrics (Miller et al., 2010); one is quasi-experimental study (Russell et al., 2015); one is a comparison study that quantified student interview data (Hanauer et al., 2012); and two are qualitative studies (Iimoto &

Frederick, 2011; Quardokus et al., 2012). The specific information is presented in the Figure 6.

The examination of authenticity also found that authentic scientific inquiry or authentic research experiences are interpreted in a variety ways. Some studies have claimed that their learning environments allow students to experience authentic research experiences, but a careful examination found that studies provided little information about the authenticity of the project (e.g., Bernard, 2011; Cakir, 2011). A few studies defined authenticity as the highest level of independence in which students design research questions and procedures (e.g., Gornally, Brickman, Hallar, & Armstrong, 2011; Nadelson et al., 2010). Some research questions are real-world problems, but with little information about its theoretical background and investigative values (e.g., Bussey et al., 2015; Campbell et al., 2012; Powell & Harmon, 2014;). In some studies, students were exposed to a natural environment and experienced the scientific inquiry steps, but did not investigated original research questions (e.g., Lustick, 2009; Nugent et al., 2008; Schussler et al., 2013). Some inquiry tasks were hands on activity, but not for answering an original research question (e.g., Bergwerff & Warners, 2007; Nugent et al., 2012; Song & Schwenz, 2013). In a few studies, students posed a hypothesis or proposal from doing literature reviews for independent research (e.g., Chung & Behan, 2010; Iimoto & Frederick, 2011). In a nutshell, in inquiry activities described in many studies, students participated in many cognitive and behavior practices that scientists perform; however, the purpose and motivation for the inquiry is to challenge students rather than make an original intellectual or creative contribution to the discipline (Auchincloss et al., 2014; Hunter et al., 2006).
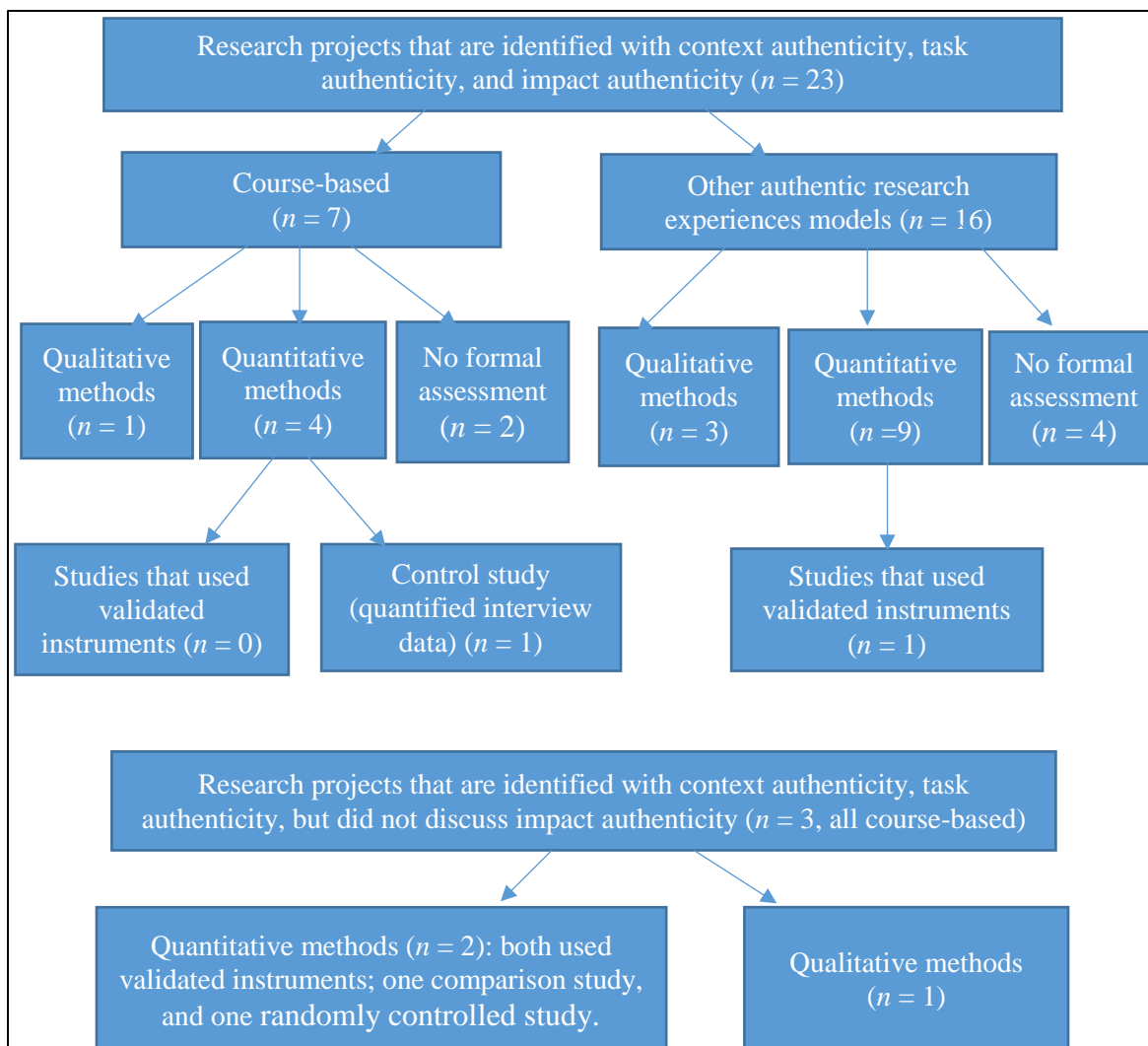
Figure 6. Information about the Authenticity of Included Studies

Evidence of the impact of undergraduate students' research experiences.

Existing studies have shown that undergraduate research experiences have a variety of

positive impacts on student learning, which is categorized and organized as: intellectual

outcomes; attitudes towards science and research; ownership and autonomy; student

confidence; student scientific inquiry and research skills; problem solving skills and

critical thinking; networking skills; collaboration and communication skills; participation

in professional meetings, journal publications, and community practice; retention and

selection in STEM related graduate education or career; view of the nature of science;

involvement in the research culture; identification as a scientist; and identification with the institution (see Table 8).

Nevertheless, a few studies have found little influence of undergraduate research experiences on student learning outcomes. One study (Behar-Horenstein & Johnson, 2010) found that the instructional design of presenting students an overview of faculty's current research topics and providing students with opportunities to interview four professors to learn more about their research and to write a report, was not engaging to students, lacked student input and participation, did not involve students in the culture of science, and made students feel doing science is inaccessible. Gormally et al. (2011) developed and implemented an inquiry-based biology laboratory curriculum in which students chose the research question, variables, and protocol and explained their results in light of other studies and theories. The authors found that students rated their experiences lower on course evaluation than students' course evaluations on traditional labs in which students followed the instruction and protocol provided by facluty. In this study, the authors defined authentic scientific inquiry as the highest level of student independence in investigation, and the inquiry-based lab did not certainly provide context authenticity, task authenticity and impact authenticity for students to experience meaningful inquiry. The study conducted by Lustick (2009) examined an inquiry-based course in which the class investigated the question ''How can peak autumn color in New England be determined?'' The study found that the course failed to achieve its learning goals. In this study, the author claimed that this course strategy is to provide students authentic inquiry experiences, but obviously it was problem-based learning that did not allow authentic research experiences.

Table 8

The Evidence of the Impact of Undergradaute Research Experiencs on Student Learning

Categories of student learning gains	Positive impact	No impact
Intellectual outcomes: concept/content knowledge comprehension and appliation; academic perfprmance, GPA	Bernard, 2011; Bussey et al., 2015; Cakir, 2011; Griffard & Golkowska, 2013; Hunter et al., 2006; Jansen et al., 2015; Jones et al., 2010; Thiry, Weston, Laursen, & Hunter, 2012; Lopatto, 2010; Miller et al., 2010; Nadelson et al., 2010; Nugent et al., 2012; Pacifici & Thomson, 2011; Pender et al., 2011; Russel et al., 2015; Russell, Hancock, & McCullough, 2007; Shields et al., 2010; Song & Schwenz, 2013;Vieyra, Gilmore, & Timmerman, 2011.	Gormally et al., 2011; Lustick, 2009; Nugent et al., 2008.
Scientific inquiry/research skills: posing hypothesis, oberserving, collecting and anlyzing data, interpreting data and results, presenting and communicating findings	Bergwerff & Warners, 2007; Bussey et al., 2015; Canaria et al., 2012; Chung & Behan, 2010; Davidson & Palermo, 2015; Ellis-Monaghan & Pangborn, 2013; Gilmore et al., 2015; Griffard & Golkowska, 2013; Iimoto & Frederick, 2011; Jansen et al., 2015; Lopatto, 2010; Miller et al., 2013; Nugent et al., 2008; Shanle et al., 2016; Urias et al., 2012; Wagner et al., 2010; Wilson, Howitt, Wilson, & Roberts, 2012; Zimbardi et al., 2012.	Lustick, 2009; Wilson et al., 2012
Problem solving, critical thinking skills	Iimoto & Frederick, 2011; Luckie et al., 2012.	
Networking: building relationship with faculty or senior researchers, finding new research opportunities.	Canaria et al., 2012; Hanauer & Hatfull, 2015; Hartmann et al., 2013; Naug et al., 2012; Pacifici & Thomson, 2011; Urias et al., 2012; Zhan, 2014.	
Participation in professional conferences, journal publications, and community practice	Culp & Urtel, 2013; Dillner et al., 2011; Jansen et al., 2015; Miller et al., 2013; Russell et al., 2007; Urias et al., 2012.	
Collaboration and communication skills	Woodzicha et al., 2015;	Ing et al., 2013;
Attitudes: interest, engagement, curiosity, satisfaction	Bernard, 2011; Campbell et al., 2012; Hartmann et al.,2013; Hunter et al., 2006; Jaarsma et al., 2009; Thiry, Weston, Laursen, & Hunter, 2012; Luckie et al., 2012; Miller et al., 2013; Nadelson et al., 2010; Naug et al.,2012; Nugent et al., 2008.	Behar-Horenstein & Johnson, 2010; Davidson & Palermo, 2015; Lustick, 2009.
Ownership and autonomy	Bernard, 2011; Gilmore et al., 2015; Hanauer et al., 2012; Zhan, 2014;	

Categories of student learning gains	Positive impact	No impact
Confidence: using techniques and tools, doing science, inquiry-based science teaching	Coverdale, 2002; Thiry, Weston, Laursen, & Hunter, 2012;; Nadelson et al., 2010; Nugent et al., 2012; Nugent et al., 2008; Pacifici & Thomson, 2011; Russel et al., 2015; Russell et al., 2007; Shanle et al., 2016; Vieyra, Gilmore, & Timmerman, 2011.	Lustick, 2009;
Retention and selection in STEM related graduate education or career	Adedokun, Zhang, Parker, Bessenbacher, Childress, & Burgess, 2012; Barker, 2009; Griffard & Golkowska, 2013; Harsh, Maltese, & Tai, 2011; Jones et al., 2010; Kardash & Edwards, 2012; Kendricks & Arment, 2011; Lopatto, 2004, 2007; Luckie et al.,2012; Nugent et al., 2008; Pender et al., 2010; Quardokus, Lasher-Trapp, & Riggs, 2012; Russell et al., 2007; Seymour, Hunter, Laursen, & DeAntoni , 2004; Shields et al., 2009; Vieyra et al., 2011; Wilson et al., 2012; Yaffe, Bender, & Sechrest, 2014; Zhan, 2014.	Naug et al., 2012
View of the Nature of Science	Adedokun & Burgess, 2011; Bergwerff & Warners, 2007; Chung & Behan, 2010; Griffard & Golkowska, 2013; Miller et al., 2013; Pacifici & Thomson, 2011; Woodzicka et al., 2015.	
Envolement in the research culture	Barker, 2009; Canaria et al., 2012; Dillner et al., 2011; Hunter et al., 2006; Jaarsma et al., 2009; Kardash & Edwards, 2012; Russell et al.,2007; Wilson et al., 2012; Zhan, 2014.	Behar-Horenstein & Johnson, 2010
Identification as a scientist	Barker, 2009; Hunter et al., 2006; Seymour et al., 2004; Wilson et al., 2012.	Shanle et al., 2016
Identification with the institution	Yaffe et al., 2014	

Factors that influence the impact of undergraduate research experiences.

Gender, ethnicity, the duration of research experiences, and the authenticity of research

projects are found as factors that influence the impact of undergraduate research

experiences on student learning outcomes. Undergraduate research experiences were

found especially beneficial for female and minority students. A few studies investigated

the effect of gender and ethnicity on student research experiences. Harsh et al. (2012) surveyed 4,285 practicing scientists and graduate students using a stratified random sampling approach (for gender), and interviewed 116 individuals who did not participate in the survey to examine the gender-based variations of the effect of the undergraduate research experiences. The results indicated that women had a significantly higher rate at identifying undergraduate research experiences as a primary reason for entering graduate school than their male counterparts. The findings suggested the long-term efficacy of undergraduate research experiences as a gateway for women interested in STEM careers and provided support in justifying research program and initiatives for women in traditionally male-dominated fields.

Another study conducted by Taraban and Logue (2012) surveyed 353 female students and 244 male students, and the results showed that male students achieved higher scores on cognitive factors associated with benefits from doing research. Students with below-average GPAs and students with average or below-average participation in research showed a decline in research benefits as they moved through their college years.

Overall, these findings showed that all students do not benefit from doing research and that the means to achieving the ideological goal of involving all students in research may vary across disciplines. There is a need for more attention to student differences as they apply to research participation, including academic ability, gender, and college level, and to the academic resources and practices that more inclusively and effectively involve students in research. Kim, Fann, and Misa-Escalante (2011) explored programmatic elements that promote gender equity and identified specific mechanisms in supporting and encouraging women to persist in computer science and engineering fields.

Analysis of data collected from surveying 117 NSF funded REU nationwide programs and 20 follow up interviews indicated that female students benefits most from participating in undergraduate research experiences that have a critical mass of female students, and even more when supportive role models are involved. Strong research experiences that involve students in the research community to investigate real-world questions are beneficial to both women and men. Gender-focused activities are most beneficial when they are presented naturally. An informal assessment conducted by Vieyra et al. (2011) found that requiring undergraduate research helped engage African-American females in STEM related fields. Pender et al. (2011) examined the effects of undergraduate research on minorities' learning outcomes by controlling for a variety of background, academic and family characteristics. Results showed that the impact of summer research experiences on academic outcomes and the retention in STEM of minorities is vital.

Jones et al. (2010) examined the association between timing and duration of undergraduate research participation and college retention and performance in the biological sciences using longitudinal data of biology majors at UC Davis. The results showed that there were no significant differences between underrepresented minorities and Asian and White students in the association between research participation and graduation outcomes, but non-Philipino underrepresented minorities had lower predicted probabilities of graduation regardless of undergraduate research status. Kendricks and Arment (2011) found that research experiences especially improved minority student performance and retention rates in STEM.

Gilmore et al. (2015) examined the relationship between undergraduate research characteristics including duration, autonomy, collaboration, and motivation, and research skills skill performance in graduate school. The authors described undergraduate research experiences as a cognitive apprenticeship model that apprentices (undergraduate researchers) gained disciplinary knowledge and skills through close interaction with recognized disciplinary experts. Fifty eight graduate students' proposals were evaluated using a previously validated rubric (Timmerman, Strickland, Johnson, & Payne, 2011) for assessing scientific reasoning skills through writing. The results found that undergraduate research experience was linked to heightened graduate school performance in all research sills assessed. Duration was most strongly correlated to significant increase in research skill performance.

Vieyra, Carlson, Leaver, and Timmerman (2013) investigated student perceptions about research and found that minority females had the highest rates of misconceptions regarding the nature of research; that research was mostly conducted in the library, similar to what they do for a class. Harsh et al. (2011) conducted a study using the data from a national mixed-methods study that surveyed 4,285 respondents and interviewed 116 individuals. The findings indicated that although research experiences afford students a multitude of benefits, the exposure to genuine, authentic research was considered the most valued attribute by the majority of respondents.

The validity and reliability of instruments for assessing undergraduate research experiences. Included studies used a variety of forms of data for assessing student learning gains from participating in undergraduate research, including interviews, surveys and open-ended questions, course evaluation, grading, within project results,

university student records such as GPA, transcripts, and other standardized tests, student reflection, reflective journals. Among these, 32 studies used surveys for data collection.

Several studies (Lopatto, 2004, 2007, 2010; Miller et al., 2013; Nadelson et al., 2010; Shanle et al., 2016) used the Classroom Undergraduate Research Experiences (CURE) survey (Lopatto & Tobias, 2010) and The Survey of Undergraduate Research Experiences (SURE) (Lopatto, 2004). Yet these surveys are limited as a measure of the nature and outcomes of undergraduate research experiences because the critical information about the reliability and validity of the surveys is missing (Auchincloss et al., 2014). Most studies used researcher-designed surveys or existing surveys, but provided little information regarding the validity and reliability of the surveys used in these studies (e.g., Canaria et al., 2012; Davidson & Palermo, 2015; Harsh, Maltese, & Tai, 2012; Nugent, Kunz, Levy, Harwood & Carlson, 2008; Kim et al., 2011; Lustick, 2009; Naug et al., 2012; Nugent et al., 2008; Pacifici & Thomson, 2011; Powell & Harmon, 2014; Urias et al., 2012; Wagner et al., 2010; Willis et al., 2013; Wilson et al., 2012; Yaffe et al., 2014; Zhan, 2014).

A few studies used surveys designed by researchers and provided information about reliability (Hartmann et al., 2013; Ing et al., 2013; Pacifici & Thomson, 2011; Woodzicha et al, 2015). Surveys used in two studies (Hartmann et al., 2013; Ing et al., 2011) are concerned with very low reliability coefficients (lower than .50). One study (Thiry et al., 2012) used a survey entitled Undergraduate Research Student Self-Assessment (URSSA), which is an online survey instrument for use in evaluating student outcomes of undergraduate research experiences in the sciences. Two articles described the development and validation of the URSSA (Hunter et al., 2009; Weston & Laursen,

2015); however, the information about the validity was of concern with the model fit statistics for the four-factor model of RMSEA = 0.064, CFI = 0.76, and chi square/*df* = 3.0. Russell et al. (2007) analyzed the data collected from a nationwide evaluation of undergraduate research opportunities (UROs); however, the validity and reliability of the instrument is not found based on the references the authors provided and an online search.

Studies that used validated instruments are presented in the Table 9. Jaarsma and colleagues (2009) conducted confirmatory factor analysis using Multiple Group Method (MGM) to validate an existed instrument, which showed that the new 5-factor model fitted the data. Hanauer, Frederick, Fotinakes, and Strobel (2012) developed and validate a simple survey instrument to measure student conversational networking by conducting an exploratory factor analysis and evaluating internal consistency using Cronbach's alpha ($\alpha = 0.88$).

Table 9

Validated Instruments Used in Included Studies

Instrument	User	Developer
Geological Time Aptitude Test	Nugent et al., 2012	GeoTAT, Dodick & Orion, 2003
A rubric for assessing scientific reasoning skills through writing	Gilmore et al., 2015	Timmerman et al., 2011
Principles of Scientific Inquiry-Student (PSI-S) surveys	Campbell et al., 2012	Campbell, Abd-Hamid, & Chapman, 2010
Views on Science-Technology-Society report (VOSTS)	Schussler et al., 2013	Aikenhead & Ryan, 1992.
Views of Nature of Science, Form B, questionnaire	Schussler et al., 2013	Bell, Blair, Grawford & Lederman, 2003; VNOS-B;

A survey instrument to measure   Hanauer et al., 2012   Hanauer et al., 2012

student conversational networking

The examination of the instruments used in the included studies indicates that validity related problems are a weakness of many included studies. Only a small portion of included studies provided the information about the reliability and validity of the instruments. The validation of the instruments used in assessment of undergraduate research experiences is mainly reported as the internal consistency coefficients using Cronbach's alpha, or factor loadings of items on latent factors. Few studies discussed the confirmation of the factorial structure and the construct validity of the instruments including discriminant validity and convergent validity.

Conclusion

This systematic literature review of the impact of undergraduate research experiences has attempted to synthesize understandings of the designs, implementation, and evaluation of undergraduate research experiences that can contribute to better undergraduate research experiences and assessments.

The analysis of 73 studies indicates that involving undergraduates in scientific research has positive impact on student learning outcomes. However, few studies investigate the impact of undergraduate student research experiences on student literacy skills, which is an important goal in science education (NRC, 1996). Integrating research experiences into undergraduate science curriculum has been increasingly adopted as an approach to expanding the opportunities of experiencing scientific inquiry to large group of students. Nevertheless, studies defined authentic research or authentic scientific

inquiry in different ways, and the authenticity of the research experiences is not clarified and examined according to authentic learning theory.

In addition, studies that use solid design and validated methods to investigate the impact of course-based undergraduate research experiences are rare. Conclusive evidence is lacking at this point regarding the impacts of integrating authentic research in undergraduate science curricula (Campbell et al., 2010) due to the fact that few researchers in the sciences have the appropriate background and experience with program assessment (Urias et al., 2012). Most studies related to undergraduate authentic research program assessment relied on qualitative research methods or researcher-designed measures, which were not statistically validated. In contrast, several instruments have been developed based on data collected from K-12 students to assess student experience in scientific inquiry. Examples include the Principles of Scientific Inquiry-Student (PSI-S) by Campbell et al. (2010) and A Competency Scale for Learning Science: Inquiry and Communication, by Chang et al. (2011). To address the lack of validated assessment tools that specifically gauge the impact of integrating authentic research into undergraduate science curricula (Auchincloss et al., 2014), it is important to develop and validate an instrument to assess student self-reported science learning gains from involvement in authentic research.

This systematic literature review suggests a need of inferential study, especially control study that compares the impact of research-based lab and traditional lab on student learning outcomes. Though benefits of undergraduate research have been reported, documented correlations do not allow a strong predictive statement to be made regarding the influence of undergraduate research on student outcomes, especially those

regarding scientific literacy skills (Auchincloss et al., 2014). Empirical analyses of the outcomes of course-based undergraduate research programs are critical for improving undergraduate research experiences and for encouraging more educators in the field of undergraduate science education to adopt this instructional approach and enhance science education more generally (Sadler & McKinney, 2010).

The study presented in this dissertation is an endeavor to bridge the research gap in the assessment of course based undergraduate research experiences. The first part of the study in this dissertation is a psychometric analysis that is used to develop and validate a new instrument (Student Science Learning Gains Survey) for assessing the undergraduate science curriculum that integrates authentic research. The second part is a predictive study that uses path analysis to investigate the predictive influence of student authentic scientific inquiry experiences on student learning outcomes. For the second research question, the level of student authentic scientific inquiry experience, which is indicated as the number of authentic research courses a student took, is the predictive variable. Student scientific literacy skills is the dependent variable. Student interest, attitudes, tool and technique skills, and communication ability were mediating variables.

Chapter 4

Undergraduate Authentic Research Project Context

The context of the presented study in this dissertation is a GENI web-based model that integrates authentic scientific research projects into undergraduate science curricula. Compared to traditional science teaching approaches that use passively repeated pre-designed experiments which report known results and where students rarely have the chance to design, practice, and discuss science as a process of inquiry, involvement in authentic research helps students deepen their understanding of the complex nature of science and enhance student inquiry competency (Bruner, 1960; Hume, 2009; Tytler, 2007).

By its nature, integrating authentic research into science curricula requires higher-level resources and instructional strategies (Hodson, 1996). Founded on situated learning theory and practical experience, the GENI web-based model of integrating authentic scientific inquiry into science curricula has been developed and implemented to teach authentic scientific inquiry to college students. A flexible online platform and database entitled Guiding Education through Novel Investigation (GENI-SCIENCE) was developed to facilitate and coordinate authentic project-based classroom research. GENI-SCIENCE (http://geni-science.org/) hosts diverse research projects across many disciplines that currently range from genomics to physical biochemistry.

One of the most important requirements to allow students to be involved in authentic inquiry in a classroom setting is to provide appropriate and accessible research questions that are under active investigation by scientists. Compared to professional scientists, students have a less sophisticated level of knowledge, experience, attitude, and

scientific reasoning (Lee & Songer, 2003), and students' inquiry is more constrained due by limited time and resources (Edelson, 1998). Research projects integrated into a particular curriculum should not only align with content knowledge and skills that are targeted in a certain course, but also with students' existing knowledge base, scientific thinking skills and course resources. When integrating authentic research projects, it is necessary to determine if the complexity of the projects that professional scientists are pursuing is beyond student knowledge, experience, and scientific reasoning. The development of tractable research projects for undergraduate students to participate in is a major barrier that keeps instructors from teaching authentic scientific inquiry in the classroom.

After many years of implementing authentic research in their classrooms, researchers in this study have designed and developed a series of research projects that are adapted for undergraduate inquirers. Examples include the annotation of bacterial genomes using GENI-ACT and readily available online tools (over 18 complete bacterial genomes under investigation by the group), examination of gene function in both eukaryotic (Caenorhabditis elegans) and multiple prokaryotic systems, genome sequencing and closure, protein purification and binding interactions in the human immune system and functional complementation of duplicated amino acid biosynthetic genes in various bacteria. These projects are shared on the GENI website with instructors who desire to adopt collaborative authentic scientific inquiry in their classrooms.

During the course of our evaluation, instructor A successfully integrated instructor B's research project in his course in order to identify challenges faced by instructors with limited personal experience in the experimental system under

investigation. A systematic assessment demonstrated that the positive impact of this model on student learning is replicable. The positive outcomes suggest that this model can work with instructors who do not have direct research experience with the model system under investigation.

Another key component of successfully implementing authentic research in a course is providing appropriate resources and guidance. According to Vygotsky's theory of the 'Zone of Proximal Development', teaching and learning is designed to close the gap between skills the student can develop without assistance, and the potential level of proficiency that the student can reach with guidance (Vygotsky, 1978). Unlike scientists, as novice inquirers students require a lot of guidance. Authentic research projects are open-ended investigations and multi-step inquiry tasks, so instructors need to provide scaffolding to address the lack of subject matter knowledge, technical expertise, understanding of the nature of science, and motivation and commitment of novice inquirers when facing uncertainty and unknown results (Edelson, 1998).

Protocols, background information, expert tips and advice along with other easily accessible resources provide scaffolding and are critical tools supporting student investigation. The GENI website contains projects used in both lower- and upper-division courses including: general biology, genetics, molecular biology, biochemistry, and physical chemistry. Each project domain contains five tabs that organize project information: Background, Syllabus, Kit Materials, Media, Reagent and Chemical List, and Equipment List. Below the tabs sit multiple dynamic windows, which contain step-by-step protocols and fields for entering data as text or files. Each window involves a discrete task that takes, for example, one lab period to complete. Within each window are

the following four sub-windows: Introduction, Protocol, Upload Results, and View

Results. Using these, students can access the protocols and upload data during and after

lab. Protocols and physical resources have been refined by researchers in this study, and

are open to customization by other instructors who adopt the research modules shared by

this study.

       In short, as a flexible tool, the GENI website serves as a platform to store and

share well-designed research modules, protocols and resources; to facilitate

communication and cooperation among instructors; to provide a learning management

system for students to access materials and resources, submit data and results, share

research progress and results, and to store and collect data that can be used by

geographically distant researchers.

Chapter 5

Development and Validation of SSLG

This chapter presents the study of developing and validating a new instrument for assessing the impact of course-based undergraduate research experiences. For the purpose of assessing the effects of undergraduate science curricula that integrates authentic scientific inquiry, assessment strategies should focus on key features of authentic scientific inquiry in the community of scientists, rather than merely focusing on knowledge presented in textbooks (Chinn & Malhotra, 2002; Edelson, 1998; Hanauer et al., 2009). In order to assess the nature and outcomes of such undergraduate program, it is necessary to review the paradigms essential to the scientific enterprise related to key features of authentic scientific inquiry (Kuhn, 1996). Only then can undergraduate science programs be accurately evaluated for their alignment with proposed educational objectives for authentic inquiry.

Paradigms of Authentic Scientific Inquiry

Kuhn (1996) defined scientific paradigms as "universally recognized scientific achievements that, for a time, provide model problems and solutions for a community of researchers" (p. 12). Scientific paradigms determine what is to be observed and scrutinized, the ways in which questions are to be structured, and how results are interpreted. The main paradigms that guide scientific practice and greatly influence science learning are reductionism and systems science (Hume, 2009).

Reductionism has existed since Descartes and the Renaissance and is rooted in the assumption that complex problems can be solved by breaking them down into smaller, simpler, and more tractable elementary units. Causal factors inherent in reductionism are

identified experimentally or by deductive reasoning derived predominantly from mathematical models (Bonaccorsi, 2010).

Systems science, which emerged in the 1950s, views the scientific process as a holistic system focusing on the relationships and interactions between its various parts. The emergence of systems science is linked to the development of transdisciplinary, which seeks to reduce boundaries within academic disciplines to more effectively address complex theoretical and practical problems that cannot be resolved by a single discipline alone (Hieronymi, 2013). Within this paradigm, an emergent pattern arises as a consequence of interactions among components or subsystems, instead of as the result of single linear-causal relationships. Systems biology is representative of a systems science approach, which integrates multiple disciplines including molecular biology, biochemistry and biophysics into a more complex and useful approach to problem solving (Fang & Casadevall, 2011; Somerville et al., 2004).

Social and cultural features of science are outlined and valued within the systems science paradigm. Communication among the science community, interactions between science and non-science institutes, integration and expansion of technology and scientists' tacit knowledge of the scientific process are all considered important contributors to identify meaningful research goals and to choose appropriate methods (Hieronymi, 2013; Hodson, 1992). Scientists highlight the importance of creativity and critical thinking in problem solving through authentic investigation (Wong & Hodson, 2009). As Kuhn (1996) claimed, changing perspectives, mental models, and methodology all lead to small scientific revolutions.

The systems science approach is claimed as "…a counter-current to the increasing fractionation of science into highly specialized branches resulting in a breakdown of communication between the specialists" (Rapoport, 1986, preface). Scholars have argued that systems science is not truly opposed to, but complementary to reductionism (Ahn, Tewari, Chi-Sang, & Phillips, 2006; Fang, 2011). Bonaccorsi (2010) even suggested that complex systems are the "result of internal dynamics epistemic of science", and interdisciplinary and transdisciplinary are the "result of application of reductionism strategies to complex multi-layered system" (p. 381). Rather than choosing one approach over the other during a scientific investigation, researchers should consider the limitations of both paradigms and treat them as interdependent and complementary (Ahn et al., 2006; Fang, 2011). Reductionism depicts the system as a collection of static components but disregards the dynamic interaction between components. While the systems science approach integrates contextual information, it is not readily applied to investigate causal factors due to the large number of confounding variables.

Features of Authentic Scientific Inquiry and Educational Objectives

The preceding overview of the two paradigms of scientific practice demonstrates that authentic scientific inquiry in educational settings involves not only a framework of actions and methods shared in the scientific community, but also that scientific reasoning and intuitive knowledge of the process of science (Hume, 2009; Reiser, Radinsky, Edelson, Gomez, & Marshall, 2001), attitudes of uncertainty and commitment (Edelson, 1998), communication within the scientific community and interactions between science and non-science institutions are transferrable in different contexts.

Despite differences in how the process of science works between these two paradigms, there is agreement in the literature about core features of the scientific process which include "abilities related to identifying investigable questions, designing investigations, obtaining evidence, interpreting evidence in terms of the question addressed in the inquiry and communicating the investigation process" (Harlen, 1999, p. 129). Peer review is regarded as the gold standard for evaluating scientific inquiry across paradigms (Popper, 1959). When scientists publish the details of their research, both techniques and results of the study are subject to other scientists' critical re-examination.

In addition, uncertainty, commitment and persistence to overcome challenges and frustrations are key features of authentic scientific inquiry across paradigms (Edelson, 1998) and are therefore critical for the design of undergraduate science curricula. These features of authentic scientific inquiry are best learned by engaging in authentic research in a laboratory or field setting (Wong & Hodson, 2009). For example, Hodson (1992) argued that scientists' personal theoretical constructs and tacit knowledge of how to do science only comes with the experience of doing science as a holistic investigation in many different contexts.

The theoretical framework of this study establishes a foundation to determine which educational objectives will be most useful in assessing students' science learning gains from performing authentic research. Due to the encouragement and support of programs such as the National Science Foundation (NSF) Research Experiences for Undergraduates (REU), undergraduate students have been involved in authentic research based learning environment. Integrating authentic research in science curricula empowers learners to meet educational objectives allowing the acquisition of "a body of

scientific knowledge that is integrated with an understanding of science knowledge, attitudes, tools, techniques, and social interaction" (Edelson, 1998, p. 320).

Correspondingly, the strategies to assess student learning gains from science curricula integrated with authentic research should address each of the following educational aspects: scientific knowledge, tools, techniques, attitudes, and social interaction. The instrument developed in this study is focused on assessing the following constructs: (1) student understanding of core concepts contextualized in authentic research; (2) student scientific inquiry skills in terms of techniques, methods, and communication; (3) student self-efficacy and attitudes toward science; and (4) student theoretical and procedural knowledge as an indicator of conceptual understanding and ability to utilize the processes of science..

Development of the Instrument

An instrument, the Student Science Learning Gains Survey (SSLG), was developed to assess students' self-reported learning gains from participation in authentic research as part of the undergraduate science curriculum. The development of the SSLG followed basic steps in survey development, including formulating the study objectives, forming the survey items, grouping items, pretesting the questionnaire with expert evaluators, pilot testing the instrument with a sample population, analyzing data for validity, main testing from a sample study population, and statistical item analyses. All statistical analyses were conducted using the Statistical Package for Social Science (SPSS) version 22 for Windows.

Instrument constructs and item formulations. Survey questions were formulated based on theoretical frameworks and previous related studies (Student

Assessment of Their Learning Gains [SALG], by Seymour, Wiese, Hunter, & Daffinrud, 2000; Competence Scale for Learning Science: Inquiry and Communication, by Chang et al., 2011; Student Interests Upon Entrance into and Perception Upon Exit from Research Experience for Undergraduate Program, by Urias et al., 2012; and General Self-Efficacy Scale (GSE) by Schwarzer & Jerusalem, 1995). The survey incorporated instrument constructs: 1) concept understanding, 2) scientific inquiry skills, 3) self-efficacy and attitudes, and 4) transferability of theoretical and procedural knowledge. These constructs were represented by 63 survey questions using a Likert scale ranging from 1 to 5, with 5 being the highest. Informal interviews with instructors and students in two authentic research courses were also conducted.

To select and refine survey items, and to identify the relatedness and discrimination of the identified constructs, two formal meetings with expert evaluators, and multiple meetings with project leaders were organized. The expert evaluators comprised eight undergraduate science instructors from four higher education institutions who had significant experience teaching authentic research courses over a five to 15 year period. For example, one instructor highlighted troubleshooting and technique inquisition as a significant science learning gain in an authentic research course.

Another instructor emphasized students' ability to transfer knowledge and reasoning gained from one course to other courses or situations. As a result of these efforts, 31 items were selected and revised to fit into defined categories. Pilot testing data collected from 60 students in two authentic research courses, along with student feedback, were used to modify ambiguously worded items. This process achieved content validity of the SSLG, demonstrating that the instrument addressed the outcomes that it was

intended to measure, and the face validity of the SSLG, demonstrating that the items are clearly verbalized and understood by the participants.

Exploratory Factor Analysis. Exploratory Factor Analysis (EFA) is a widely used statistical technique in constructing instrument to measure underlying variables in social science (Costello & Osborne, 2005). This technique is used to understand the latent structure of manifest variables and to identify groups of variables so as to reduce a data set to a more manageable size while retaining as much information as possible (Field, 2009). After the initial instrument was finalized, the SSLG was administrated to 222 college students from three universities majoring in the sciences.

Participants. Participants were enrolled in science courses, and most were enrolled in authentic research courses. There were 64 participants from a college in the Midwestern United States, 38 participants from a college on the East Coast of the United States, and 120 participants from a college in the Northwestern United States. Of the participants, 131 were female students and 91 were male students. One participant was a freshman, 69 were sophomores, 100 juniors, and 51 were senior students. Except for two participants who reported that they were majoring in computer science, the remaining participants reported majors in biology or chemistry.

Factor extraction and item selection. A principle axis factoring (PAF) analysis was conducted in this study because PAF is a correlation-focused approach seeking to reproduce the inter-correlations among variables and is generally used when the research purpose is detecting data structure (i.e., latent constructs or factors) or causal modeling. Variables in the SSLG were theoretically related in design so that a factor analysis (principle axis factoring) was conducted on the 31 items with rotation of Varimax

(orthogonal) and Oblimin (oblique) respectively at the first. The factor correlation matrix

showed those factors extracted are related to each other (Table 10).

Table 10

The factor correlation matrix of SSLG Factor Correlation Matrix

Factor	1	2	3	4	5	6
1	1.000	.346	.280	-.194	.427	-.561
2	.346	1.000	.295	-.313	.227	-.384
3	.280	.295	1.000	-.271	.416	-.385
4	-.194	-.313	-.271	1.000	-.220	.183
5	.427	.227	.416	-.220	1.000	-.491
6	-.561	-.384	-.385	.183	-.491	1.000

Extraction Method: Principal Axis Factoring.

Rotation Method: Oblimin with Kaiser Normalization.

Therefore, the results of the orthogonal rotation should not be trusted and the

obliquely rotated solution is more meaningful. The Kaiser-Meyer-Olkin measure verified

the sampling adequacy for the analysis, *KMO =*. 904 ("superb" according to Field, 2009),

and all *KMO* values for individual items were >.84 which were well above the acceptable

limit of .5 (Field, 2009). The determinant of the *R*-matrix is .000165, which is greater

than .00001. Therefore, the variables were inter-correlated with each other in a desirable

way for factor analysis.

Bartlett's test of sphericity χ^2 (465) = 3373.50, *p* < .001, indicated that

correlations between items were sufficiently large for PAF. Six factors were extracted

with eigenvalues over Kaiser's criterion of 1 and in combination explained 56.98% of the

variance. Due to the scree plot showed there were four factors from "cliff", the data was

run three more times, setting the number of factors extracted at four, five, and seven.

Comparing the item loading tables, the six-factor model had the best fit to the data, which

had the lowest cross loadings, and no factor with fewer than three items (Costello &

Osborne, 2005).

Therefore, given the sample size, the convergence of the scree plot, and Kaiser's

criterion of 1, six factors were retained in the final analysis. The reproduced correlation

matrix provides the information about the fit of the model to the observed data. For these

data, the footnote summary showed there were 63 (13%) non redundant residuals with

absolute values greater than .05. The percentage of 13%, which is smaller than 50%

(Field, 2009), indicated this model was a goof fit of the data.

Items were selected based on a series of criteria in terms of community, primary

factor loading, item cross-loadings, meaningful and useful membership to a factor,

interpretative purpose, and reliability (Stevens, 2009). The community indicates the

amount of variance in each item that can be explained by the extracted factors. Ideally,

the community of an item should be above .5, and researcher should consider either

removing an item with a community of less than .40 or adding similar items for future

research (Costello & Osborne, 2005). Based on these criteria, item Q5.1, which has a

community of .193 (Table 11) was dropped from the SSLG.

Table 11

The community of items of SSLG Communalities

Question	Initial	Extraction
Q 1.1	.602	.580
Q 1.2	.657	.648
Q 1.3	.648	.695
Q 1.4	.509	.390
Q 2.1	.500	.375
Q 2.2	.640	.591
Q 2.3	.645	.563
Q 2.4	.697	.691
Q 2.5	.648	.618
Q 2.6	.731	.802
Q 2.7	.571	.582
Q 2.8	.633	.718
Q 2.9	.674	.606
Q 2.10	.685	.613
Q 2.11	.746	.762
Q 2.12	.644	.651
Q 2.13	.660	.702
Q 2.14	.458	.371
Q 3.1	.649	.543
Q 3.2	.710	.717
Q 3.3	.582	.502
Q 3.4	.655	.583
Q 3.5	.651	.579
Q 3.6	.515	.414
Q 3.7	.626	.612
Q 4.1	.654	.575
Q 4.2	.548	.369
Q 4.3	.605	.637
Q 4.4	.583	.595
Q 5.1	.353	.193
Q 5.2	.447	.388

Extraction Method: Principal Axis Factoring.

According to the rules that Hair, Anderson, Tatham, and Black (1998) provided

for assessing the practical significance of standardized factor loadings in pattern matrix, a

cut-off of .40 is used in this study due to the sample size of around 250. Four items with a

community close to .4 were retained for their acceptable factor loadings. Three items (Q 3.3, Q 4.2 and Q 5.1) were removed because their loadings were less than .40 for each factor, whereas item Q 2.1 with factor loadings of 3.20, were kept for its theoretically meaningful membership to the factor 1. Items Q 2.11, Q 3.4, Q 3.7, and Q 4.1 were cross-loading items that loaded at .32 or higher on two factors. Item Q 4.1 was removed because the discrepancy between the primary and secondary factor loadings, .012, was not sufficiently large (Matsunaga, 2010). Items Q 2.11, Q 3.4, and Q 3.7 were retained because their primary factor loadings were greater than or around .5 (Costello & Osborne, 2005). Table 12, the structure matrix, shows the correlations between the variables and factors. Table 13, the pattern matrix, shows the factor loadings after rotation.

Table 12

Structure Matrix

	Factor					
	1	2	3	4	5	6
Q 2.13	.812	.413	.251	-.126	.444	-.566
Q 2.12	.781	.351	.239	-.272	.369	-.571
Q 2.11	.762	.384	.349	-.479	.491	-.660
Q 2.10	.668	.317	.320	-.350	.570	-.613
Q 2.1	.519	.403	.340	-.165	.410	-.437
Q 1.3	.259	.825	.303	-.247	.265	-.305
Q 1.2	.336	.794	.204	-.175	.124	-.315
Q 1.1	.212	.742	.136	-.299	.041	-.215
Q 1.4	.239	.611	.236	-.217	.246	-.323
Q 4.1	.374	.572	.396	-.566	.458	-.393
Q 4.2	.289	.415	.401	-.410	.405	-.399
Q 3.2	.164	.215	.823	-.079	.373	-.368
Q 3.1	.182	.261	.732	-.196	.282	-.317
Q 3.5	.482	.275	.692	-.262	.455	-.395
Q 3.7	.557	.396	.667	-.331	.427	-.478
Q 3.4	.431	.129	.628	-.103	.595	-.381
Q 5.2	.095	.216	.594	-.311	.216	-.199
Q 3.6	.297	.279	.561	-.409	.381	-.369
Q 3.3	.469	.508	.527	-.368	.439	-.413
Q 5.1	.287	.289	.312	-.260	.151	-.296
Q 4.3	.471	.489	.415	-.670	.417	-.401
Q 4.4	.361	.459	.553	-.593	.460	-.418
Q 2.8	.392	.234	.327	-.257	.842	-.432
Q 2.7	.333	.303	.344	-.169	.744	-.469
Q 2.9	.617	.330	.296	-.290	.642	-.605
Q 2.14	.292	.078	.366	-.164	.582	-.337
Q 2.6	.478	.333	.372	-.237	.451	-.891
Q 2.4	.558	.366	.407	.001	.387	-.785
Q 2.5	.431	.305	.336	-.267	.456	-.772
Q 2.3	.530	.319	.319	.015	.458	-.696
Q 2.2	.581	.412	.353	-.026	.477	-.687

Extraction Method: Principal Axis Factoring.
Rotation Method: Oblimin with Kaiser Normalization.

Table 13

Pattern Matrix

	Factor					
	1	2	3	4	5	6
Q2.13	.685	.152	-.046	.084	.102	-.106
Q2.12	.660	.030	-.054	-.117	-.021	-.199
Q2.11	.519	-.017	-.018	-.321	.067	-.290
Q2.10	.399	-.023	-.039	-.188	.256	-.253
Q2.1	.320	.213	.109	.039	.157	-.064
Q1.3	-.056	.834	.056	.032	.110	.053
Q1.2	.111	.799	-.008	.076	-.088	-.006
Q1.1	.007	.758	-.061	-.107	-.125	.015
Q1.4	-.047	.569	.010	-.011	.091	-.080
Q3.2	-.133	.005	.831	.157	.058	-.120
Q3.1	-.062	.051	.724	.009	-.038	-.074
Q5.2	-.101	.032	.577	-.170	-.027	-.002
Q3.5	.301	-.017	.577	-.038	.106	.048
Q3.7	.343	.083	.501	-.093	.015	-.036
Q3.4	.228	-.127	.497	.108	.369	.052
Q3.6	.037	.011	.403	-.251	.091	-.098
Q3.3	.215	.282	.298	-.130	.143	.025
Q5.1	.142	.114	.197	-.146	-.111	-.125
Q4.3	.219	.191	.105	-.511	.118	-.013
Q4.4	.034	.172	.288	-.407	.165	-.066
Q4.1	.047	.358	.072	-.370	.236	-.018
Q4.2	-.032	.208	.151	-.243	.185	-.143
Q2.8	.025	.027	-.056	-.075	.828	-.009
Q2.7	-.072	.134	-.002	.033	.691	-.125
Q2.14	.046	-.127	.157	-.035	.493	-.051
Q2.9	.304	.033	-.087	-.112	.399	-.238
Q2.6	-.041	-.031	.021	-.082	.006	-.900
Q2.5	-.035	-.024	-.001	-.128	.087	-.735
Q2.4	.180	.084	.158	.214	-.049	-.654
Q2.3	.183	.079	.047	.212	.135	-.518
Q2.2	.241	.177	.065	.203	.142	-.426

Extraction Method: Principal Axis Factoring.
Rotation Method: Oblimin with Kaiser Normalization.[a]
a. Rotation converged in 16 iterations.

Factor 1, with an eigenvalue of 11.91, accounted for 38.42% of the overall

variance. Five items (Q2.13, Q2.12, Q2.11, Q2.10, and Q2.1) clustered on this factor

representing scientific communication. Factor 2, with an eigenvalue of 2.37, accounted for 7.64% of the overall variance. Four items (Q1.3, Q1.2, Q1.1, and Q1.4) clustered on this factor, the same as construct 1 in the initial instrument which represents student understanding of main concepts. Factor 3, with an eigenvalue of 2.28 accounted for 7.34% variance. Seven items (Q3.2, Q3.1, Q5.2, Q3.5, Q3.7, Q3.4, and Q3.6) clustered on this factor. This cluster represents self-efficacy and attitude as originally designed. Two items (Q4.3 and Q4.4) clustered with factor 4. This cluster is the same as construct 4 in the initial instrument except item 4.1 and 4.2 that were not retained. Factor 4 represents knowledge transference. Factor 5, with an eigenvalue of 1.17, accounted for 3.78% of overall variance. Four items (Q2.8, Q2.7, Q2.14 and Q2.9) clustered on factor 5 representing experiment operation skills. Factor 6, with an eigenvalue of 1.04, accounted for 3.36% of overall variance. Five items (Q2.6, Q2.5, Q2.4, Q2.3, and Q2.2) clustered with this factor that reflects planning and modifying investigation.

Items clustering on factor 1, factor 5 and factor 6 were grouped with construct 2 of scientific inquiry skills in the initial instrument, but the factor analysis grouped them into three factors. These 14 items refer to a range of scientific activity from making observation and posing hypotheses to presenting results and writing academic papers. These actions indicate multiple facets of inquiry and are integral in a holistic investigation (NRC, 1996). It is contentious whether algebraic factors represent real-world dimensions, and theoretical ground should be taken into account for refining the instrument structure (Field, 2009). Considering the content validity and ambiguous relatedness between items in these three factors, after further discussion and expert judge, these 14 items were regrouped into one factor named scientific inquiry skills based on

theoretical grounds and expert evaluation. The new SSLG instrument was finalized with 27 items categorized into the following four constructs: self-efficacy and attitude represented by seven items; concept understanding by four items; scientific inquiry skills by 14 items; and transference of knowledge by two items.

Reliability. Another basic goal of instrument development is to attain maximal reliability. Results showed that the overall *Cronbach's α = .94*, indicated a high level of internal consistency for the scale. The Cronbach's Alpha for subscales of six factors was as follows: .85, .82, .86, .80, .78, and .87. The Cronbach's Alpha for the subscale of the new factor which is a combination of factor 1, factor 5 and factor 6 extracted from the factor analysis, was .92. Reliability coefficients were all close to or above .80, which showed good internal consistency (Field, 2009). Values of corrected item – total correlations were all above .30 in all subscales. Therefore, values of Cronbach's Alpha in terms of the scale and six subscales indicated a fairly good level of internal consistency within this specific sample. The values of Corrected Item – Total Correlations were all above .30 in the two subscales. Therefore, the values of Cronbach's Alpha, in terms of the scale and six subscales, indicated a high level of internal consistency within this specific sample. Additionally, the Cronbach's Alpha if the item is deleted showed that removing any item would not improve the overall reliability of both the scale and subscales. Therefore all 27 items were retained for the next stage of the instrument development.

Confirmatory Factor Analysis. After the underlying structure of the SSLG was identified, a confirmative factor analysis (CFA) was used to verify the number of underlying dimensions of the SSLG instrument that has been established on prior EFA; to

identify the pattern of item-factor relationships; find the construct validity and the

reliability of SSLG; and to revise and refine the factorial structure of the SSLG (Floyd &

Widaman, 1995; Hernandez, 2010).

Participants. The SSLG survey explored from the EFA was administrated to 401

college students who were involved in course-based authentic research in four

universities in the Midwest and the Northwestern United States. Of the participants, 124

were male and 277 were female. Four students were freshman, 99 were sophomores, 155

were juniors, and 143 were seniors. The ethnic composition of the sample included 284

Caucasian, 66 Asian, 11 African-American, 11 mixed, and two Hispanic students.

Model structure and model fit. A confirmatory factor analysis (CFA), a type of

structural equation model (SEM), was conducted using AMOS 22 based on 401

responses. A six-factor model that was developed from the prior EFA, and a four-factor

model that regrouped three factors into one factor of Inquiry Skills were examined first.

Missing data was handled by list-wise deletion. The results showed that the goodness of

fit of the six-factor model and four-factor model was poor due to two main issues. First,

the loadings of certain variables on the factor of Self-efficacy and Attitude were low.

Since the factor of Self-efficacy and Attitude included variables pertaining to confidence

and interest, this factor was separated into two factors named as Confidence and Interest

respectively. The other issue was the high construct inter-correlations among four factors

(three factors were related to Inquiry skills and one to Transference). These four factors

were influenced by a broader factor that was named Inquiry Skills. Therefore, a second

order CFA was conducted to examine a four-factor model comprised of Interest,

Confidence, Concept Understanding, and Inquiry Skills. Whereas the second order factor

model had an acceptable goodness of fit, the indices of construct validity indicated that the factor of Confidence had discriminative validity and convergent validity concerns. An examination of the model estimates found that, the high construct inter-correlations between the factor of Confidence and the factor of Inquiry Skills caused the convergent validity and discriminant validity concerns (Farrell, 2010). After modification, a second-order factor analysis was conducted to verify a three-factor model.

The six-factor, four-factor, higher order four-factor model, and higher order three-factor model are shown in Figures 7-10 respectively. Table 14 displays the indices of model fit, the acceptance values, and the goodness of fit of the four models tested in this study. As suggested by Hu and Bentler (1999), three fit indices were mainly used to evaluate the fit of the model to the data. The root-mean-square error of approximation (RMSEA; a value of .06 or less suggests adequate model fit), was used as the index of absolute fit. The comparative fit index (CFI; a value of .95 or greater suggests adequate model fit) was used as the index of incremental fit. The value of Chi Square / df (a value smaller than 5.0 suggests adequate model fit). The model estimates indicated a good fitness of the second order three-factor model of the SSLG survey with the index of absolute fit EMSEA = .049, the index of incremental fit, CFI = .952, and the index of parsimonious fit, ChiSquare / df = 1.97. In conclusion, the results show that, the second order three-factor model appeared to provide the best fit to the data.
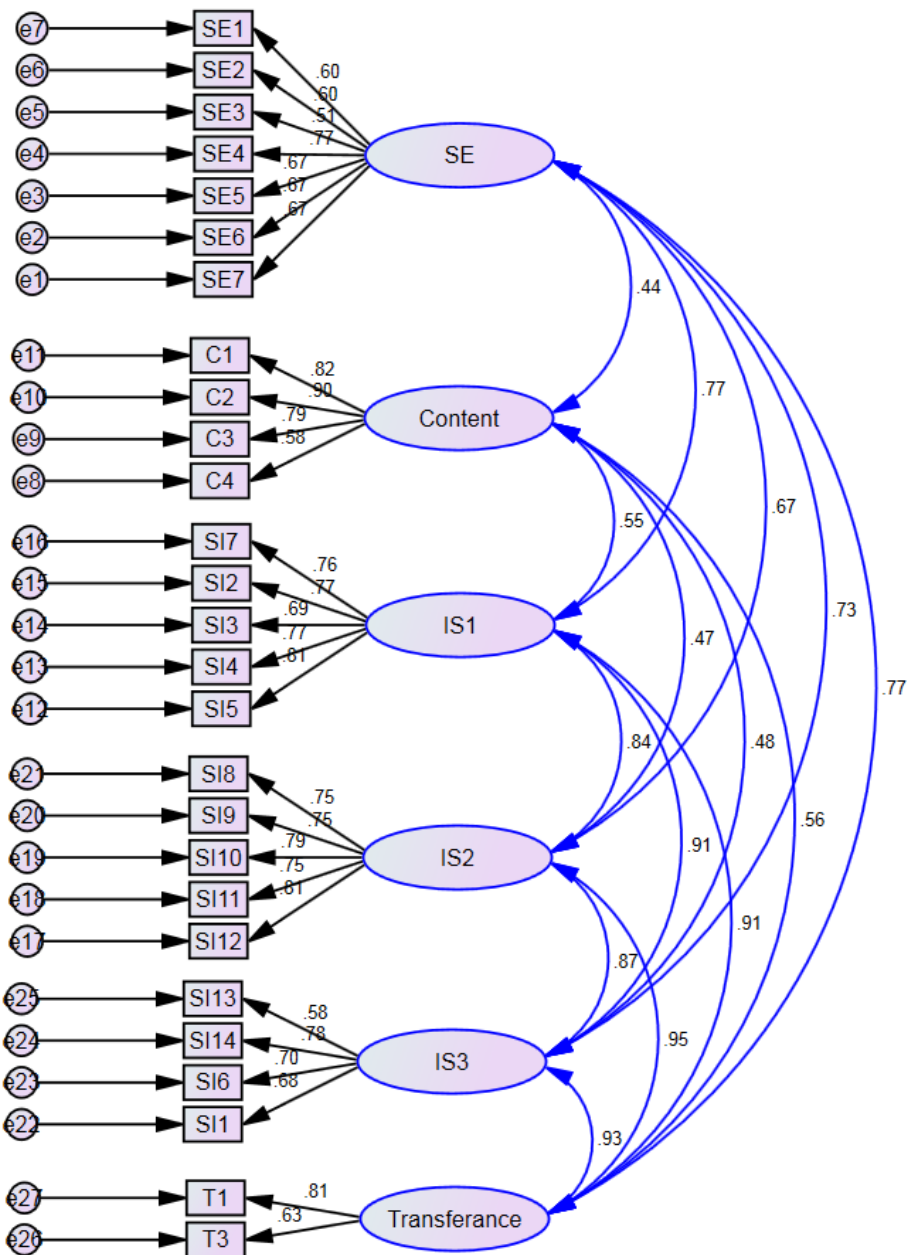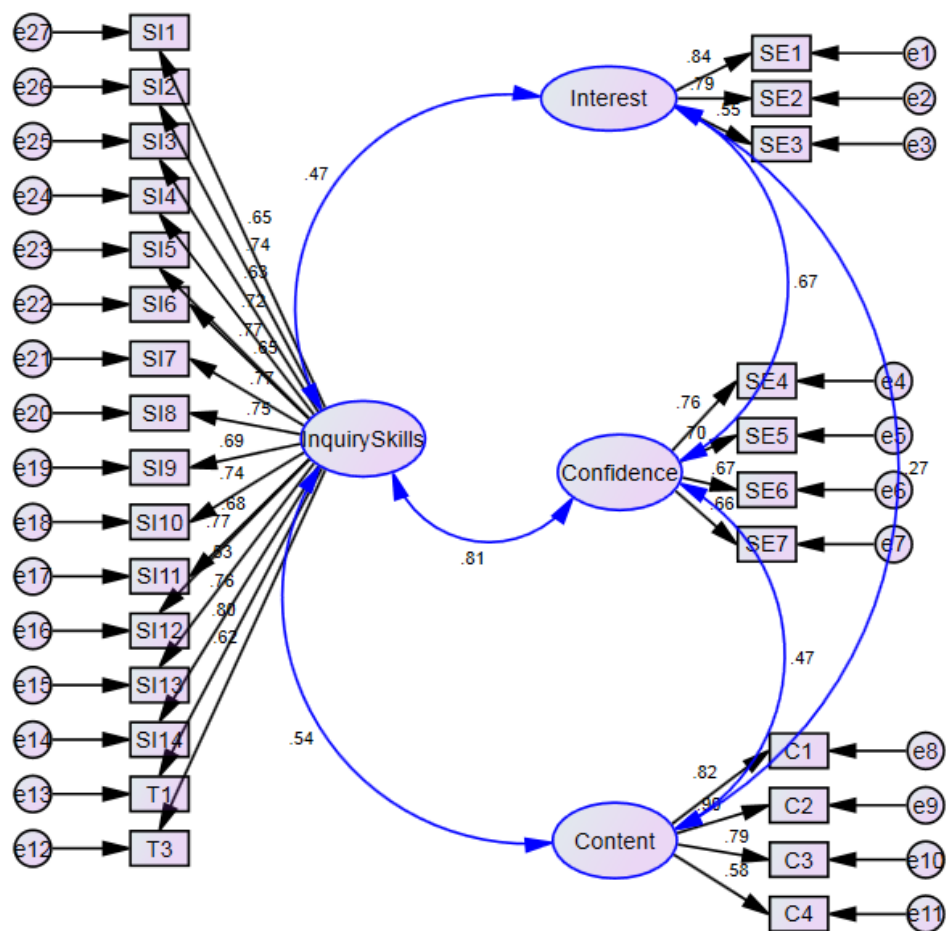
Figure 7. The Six-Factor CFA Model.
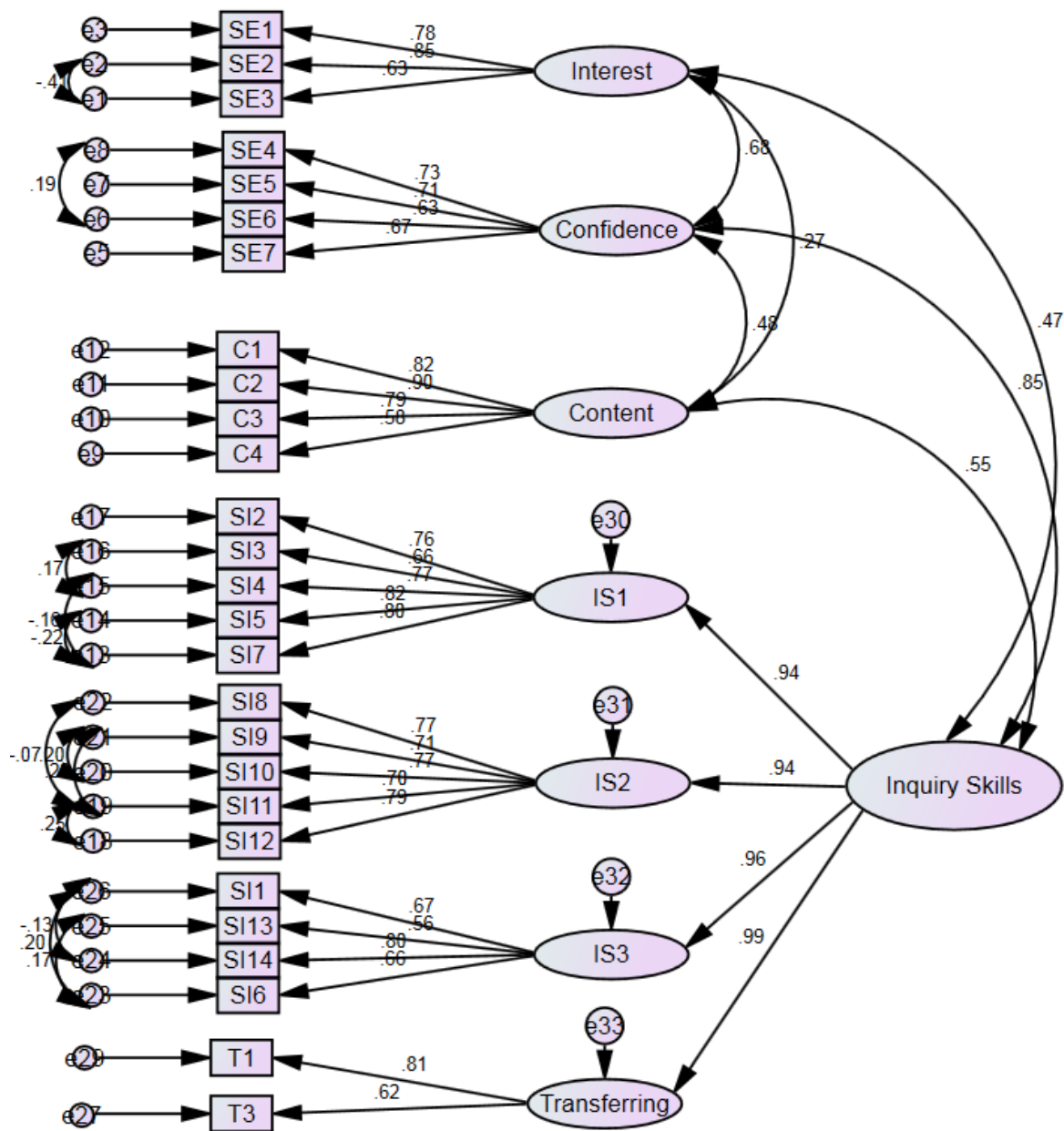
Figure 8. The Four-Factor CFA Model.

Figure 9. The Higher Order Four-Factor Model.

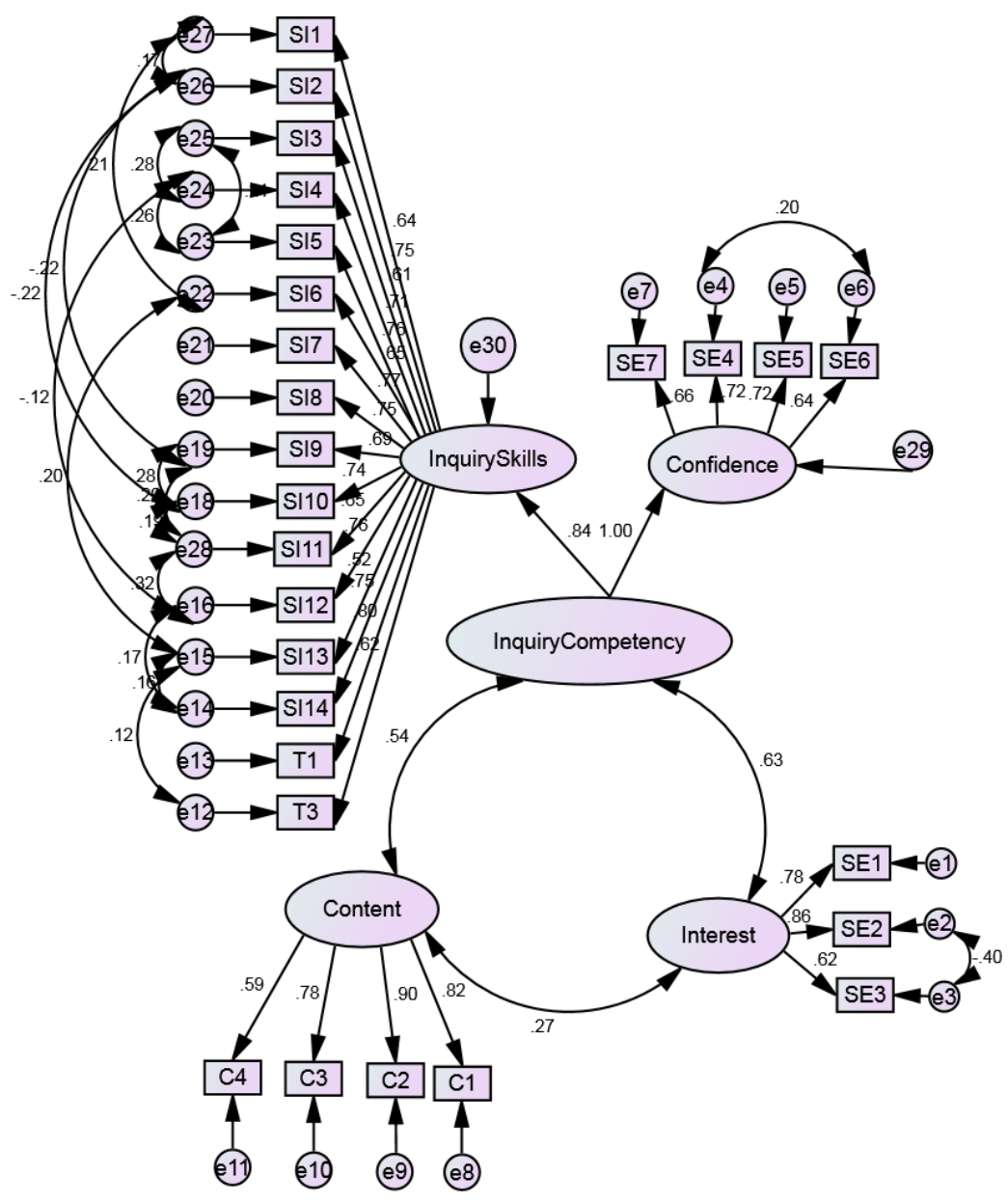*Figure 10.* The Higher Order Three-Factor Model.

Table 14

Fitness				SSLG survey			
Name of category	**Name of index**	**Level of acceptance**	**Literature**	**Six-factor model**	**Four-factor model**	**Higher oder four-factor model**	**Higher oder three-factor model**
Factor loading	Standardized regression weight	Weight > 0.6	Hair et al. (2006)	The weight of item SE3,C4 and SI13 <0.6	The weight of item SE3,C4 and SI13 <0.6	The weight of item C4 and SI13 < 0.6	The weight of item C4 and SI13 < 0.6
Absolute fit	ChiSq	P > 0.05	Wheaton et al. (1977)	.000	.000	.000	.000
	RMSEA	RMSEA < 0.08	Browne and Cudeck (1993)	.068	.067	.052	.049
	PCLOSE	PCLOSE >.05	Browne and Cudeck (1993)	.000	.000	.287	.589
	GFI	GFI > 0.9	Joreskog and Sorbom (1984)	.855	.845	.897	.902
Incremental fit	AGFI	AGFI > 0.9	Tanaka and Huba (1985)	.826	.816	.871	.877
	CFI	CFI > 0.9	Bentler (1990)	.904	.904	.946	.952
	TLI	TLI > 0.9	Bentler and Bonett (1980)	.891	.894	.937	.944
	NFI	NFI > 0.9	Bollen (1989)	.861	.859	.901	.907
Parsimonious fit	ChiSq/ df	ChiSq/ df < 5.0	Marsh and Hocevar (1985)	2.868	2.817	2.081	1.966

The Goodness of Fit of the Four Models

Construct validity and reliability of SSLG. Another goal of this study was to attain validity and reliability of SSLG survey. Content validity and face validity of SSLG survey was achieved in earlier stages of the instrument development. In this part, the focus was the construct validity and the reliability of the revised SSLG survey. Construct validity refers to the extent to which a measure adequately assesses the construct it purports to assess (Nunnally & Bernstein, 1994). Campbell and Fiske (1959) proposed

two aspects to assess the construct validity. One is convergent validity, which refers to the degree of confidence that a trait is well measured by its indicators. The other one is discriminant validity, which refers to the degree to which measures of different traits are unrelated. Confirmatory Factor Analysis (CFA) is an indispensable analytic approach for construct validation. In a confirmatory factor analysis, convergent validity and discriminant validity examine the extent to which measures of a latent variable shared their variance and how they are different from others.

In this study, convergent validity was assessed by factor loading, Average Variance Extracted (AVE), Composite Reliability (CR), internal reliability, and Discriminant Validity (Fornell & Larcker, 1981). Confirmatory Factor Analysis (CFA) was conducted to estimate factor loading of variables. A factor loading presents the level of a regression path from a latent to its indicators. In this study, all of latent variables had at least three indicators (the questionnaire item), and the value of all factor loadings was greater than .5, which was acceptable, and most factor loadings were greater than .7, which were considered as strong indicator (Hair et al., 1998). The AVE measures the level of variance captured by a construct versus the level due to measurement error. Values of AVE above .7 are considered very good, and values of .5 are acceptable (Fornell & Larcker, 1981; Zainudin, 2012). CR is another guideline to review convergent validity and the acceptable value of CR is .7 and above (Zainudin, 2012). Cronbach's alpha is a very popular coefficient to test internal reliability, and the acceptable value is above .6 (Zainudin, 2012). Discriminant validity can be assessed by comparing the amount of the variance captured by the construct (AVE) and the shared variance with

other constructs. It means the values of square root of the AVE for each construct should be greater than the correlation involving the constructs (Fornell & Larcker, 1981).

Scores on the scale of the reliability and construct validity of the higher order three-factor model of SSLG yielded good estimates. Table 15 presents the information of the construct validity and reliability of the higher order three-factor model of SSLG. The values of square root of average variance extracted (AVE) for three constructs were .76, .92, and .78, and the values of correlations between two constructs were .63, .54, and .27. The square root of AVE for each construct of the higher order three-factor model was greater than the absolute values of correlations with another construct. Therefore, the discriminant validity of this model was supported (Fornell & Larcker, 1981).

Table 15

Validity and Reliability of the Higher Order Three-Factor Model

Validity and Reliability Indices				SSLG Higher Order Three-Factor Model			
Name of Category	Name of Index	Level of Acceptance	Literature	Interest	Content	Inquiry Competency	Overall
Convergent Validity	Average Variance Extracted	AVE $\geqslant$ .5	Zainudin (2012)	.578	.61	.852	
Internal Reliability	Cronbach Alpha	$\alpha \geqslant$ .6	Zainudin (2012)	.78	.85	.94	.95
Construct Reliability	Composite Reliability	CR $\geqslant$ .6	Zainudin (2012)	.80	.86	.92	

Discussion of the Instrument Development

Integrating authentic research into undergraduate science curricula allows students to experience authentic scientific inquiry, which has been emphasized as "the central strategy for teaching science" (NRC, 1996, p. 31), to foster a deep and integral understanding of content knowledge, as well as scientific reasoning and practice (NSTA, 2007). Nevertheless, experimental evidence of the impact of authentic scientific inquiry

on student science learning is limited mainly because few valid assessment instruments exist (Auchincloss et.al., 2014). The Student Science Learning Gains instrument addresses this gap in assessing the impact of authentic scientific inquiry on student learning. Constructs and items of the SSLG instrument were formulated based on a theoretical framework that identified key features of authentic scientific inquiry from the perspective of reductionism and systems science as well as insights from scientist-educators who have been teaching authentic research courses for years.

The exploratory factor analysis indicated that the construct pattern of the SSLG was thorough and complete, and reliability was high. The confirmatory factor analysis indicated that, scores on the scale for measuring goodness fit, construct validity and reliability yielded estimates of a higher order three-factor model of SSLG with 27 items. The 27 items were categorized into the following three constructs: Interest with three items; Concept Understanding with four items; and Scientific Competency with 16 items. The brief description of items, constructs, and factor loadings of the final SSLG are shown in Table 16.

The factor of Inquiry Competency as a higher order factor includes two factors that are Confidence and Inquiry Skills. The four variables clustered in the factor of Confidence are about student self-efficacy in "overcome obstacle", "work hard and be persistent", "as an intelligent contributor", "have well-defined strategies", which reflects one key feature of the authentic scientific inquiry - attitudes (Edelson, 1998). The variables clustered in the factor of Inquiry Skills are related to skills of using tools, technique, and communication, which reflect the other two key features of authentic scientific inquiry - tools and techniques and social interactions (Edelson, 1998).

Therefore, these 16 items as indicators of inquiry competency greatly align with the features of authentic scientific inquiry. The process of modifying the model to achieve goodness of fit, construct validity and reliability manifested that CFA is a useful technique to revise and refine the factorial structure of a measurement (Floyd & Widman, 1995).

Table 16

Brief Description of Items, Latent Variables, and Factor Loadings

Item description	First-order Latent variable	Second-order latent variable	Factor loading
SE1.Think authentic scientific practice is interning.			.78
SE2. Enthusiastic in authentic inquiry.		Interest	.86
SE3. Have high expectation of learning experience.			.62
C1. Understand main concepts taught in this course.			.80
C2. Understand connections among main concepts taught in this course.		Concept Understanding	.90
C3. Link main concepts taught in this course to other courses.			.79
C4. Apply of concepts learned in daily life issue.			.59
SE4. Confidence in overcoming obstacles in scientific investigation.	Attitudes	Inquiry Competency	.72
SE5. Have well-defined problem-solving strategy.			.72
SE6. If work hard and persist, I can attain good results.			.64
SE7. See myself am intelligent contributor.			.66
SI1. Follow a scientific protocol.	Inquiry Skills	Inquiry Competency	
SI2. Pose hypothesis.			.75
SI3. Use instrumentation and lab techniques.			.61
SI4. Identify influential variables and problems.			.71
SI5. Refine experimental steps based on observation.			.76
SI6. Observe and record results.			.65
SI7. Recognize quality results.			.77
SI8. Explain and synthesize experimental results.			.75
SI9. Critically read scientific literature and information.			.69
SI10. Recognize sound argument and use evidence in augment.			.74
SI11. Write research documents in **a** discipline-appropriate style.			.65
SI12. Present and discuss data and results using techniques.	Inquiry Skills	Inquiry Competency	.76
SI13. Work effectively with others.			.52
SI14. Compare data from multiple sources.			.75
T1. Use systematic reasoning.			.80
T2. Apply scientific knowledge in real issues.			.62

Chapter 6

The Predictive Power of Authentic Research Experiences on Student Science

Learning

As science and technology innovates the world, scientific literacy becomes a

necessity for everyone (American Association for the Advancement of Science, 1989).

Though not every student will become a professional scientist, science education

familiarizes students with the natural world and with scientific concepts and processes, so

that they are able to value and apply scientific information in real world issues throughout

their lives (Hartmann, 2013). A scientifically literate person is one who is able to "know,

use, and interpret scientific explanations of the natural world; generate and evaluate

scientific evidence and explanations; understand the nature and development of scientific

knowledge; and participate productively in scientific practice and discourse" (Duschl,

Schweingruber, & Shouse, 2007, p. 2). Scientific literacy is a core goal of science

education (NRC, 1996).

Scientific inquiry competency is a critical component of scientific literacy (NRC,

1996). Higher levels of scientific inquiry skills are positively correlated with student

scientific literacy (Godek, Kaya, & Polat, 2015). Involving students in authentic scientific

inquiry processes, as a complex and contextualized enterprise, is advocated as an

instructional approach to improve both student achievement and attitudes towards science

so as to foster student scientific literacy (Anderson, 2002, Hodson, 1996).

Integrating authentic research into the undergraduate science curriculum is a form

of learner-centered active learning. Authentic scientific inquiry in science education

allows students to do science by modeling the ways in which scientists study the natural

world (Atkin & Black, 2003; Chinn & Malhotra, 2002). This activity is not about getting

expected results or right answers, but rather allows learners to investigate the natural

world in a logical and systematic fashion by proposing assumptions and interpreting and

justifying their assertions based upon evidence derived from authentic scientific work

allowing them to better understand the nature of science (Hofstein & Lunetta, 2003;

Hume, 2009). When students participate in authentic research, they approach

investigations like a scientist (NRC, 1996) working on problems that are currently

studied and debated by the community of scientists. Such problems engage students in

inquiry processes wherein student scientific knowledge is constructed and structured to

add meaning and utility. More importantly, new knowledge may be generated and

validated which could allow students to contribute valued data to the scientific

community (Hume, 2009; Reiser et al., 2001). As Edelson (1998) pointed out, the goal of

integrating authentic science research into the curriculum is to enable students to "acquire

a body of scientific knowledge that is integrated with an understanding of science

knowledge, attitudes, tools, techniques, and social interaction" (p. 320).

The value of involving undergraduate students in research is clearly recognized

(NRC, 1996). Benefits of undergraduate research have been reported, but documented

correlations do not allow a strong predictive statement to be made regarding the influence

of undergraduate research on student outcomes, especially those regarding scientific

literacy (Auchincloss et al., 2014). In this present study, path analysis was used to

examine the relationships between authentic research curriculum and student outcomes.

We examined the predictive influence of student authentic research experiences on

student interest in science, authentic scientific inquiry competency, and student scientific

literacy. In addition, we examined the predictive influence of student interest and scientific inquiry competency on student scientific literacy. Scientific competency in this study refers to three sub-categories: attitudes, tools and techniques, and communication skills.

Method and Data Sources

The Test of Scientific Literacy Skills (TOSLS) developed and validated by Gormally, Brickman, and Lutz (2012) was used to measure student scientific literacy levels. Subscales in the Student Science Learning Gains (SSLG), developed by researchers in this study, were used to measure student interest in authentic scientific practice as well as three features of authentic scientific inquiry competency: student attitudes, tools and techniques and communication skills (Edelson, 1998). Items and latent factors selected from the SSLG are presented in Table 17.

Data for this study were collected from 451 undergraduate students before they took science courses that integrated authentic research in four universities located in the west, northwest, and mid-west regions of the United States. Demographic information is presented in the Table 18. The number of authentic research science courses a student had taken was collected as an indicator of the level of research experience. A latent structural equation model was conducted using AMOS 22.

Table 17

Items and Latent Variables from the SSLG survey Included in the Path Analysis

Item	Description	Latent variable
AII	Think authentic scientific practice is interesting.	Interest
EAI	Enthusiastic in authentic inquiry.	
Exp	Have high expectation of learning experience.	
COB	Confidence in overcoming obstacles in scientific investigation.	Attitudes
SPS	Have well-defined problem-solving strategy.	
Comtt	If work hard and persist, I can attain good results.	
IntCon	See myself am intelligent contributor.	
FSP	Follow a scientific protocol.	Tools and techniques skills
PoHy	Pose hypothesis.	
UILT	Use instrumentation and lab techniques.	
IVaP	Identify influential variables and problems.	
RePro	Refine experimental steps based on observation.	
ObRD	Observe and record results.	
RQRD	Recognize quality results.	
ESER	Explain and synthesize experimental results.	
CRSL	Critically read scientific literature and information.	
ComDat	Compare data from multiple sources.	
SyReas	Use systematic reasoning.	
ASRI	Apply scientific knowledge in real issues.	
RSSA	Recognize sound argument and use evidence in augment.	Communication skills
WRP	Write research documents in **a** discipline-appropriate style.	
CRR	Present and discuss data and results using techniques.	
WwO	Work effectively with others.	

Table 18

Demographic Information

Gender (*n*)		School Year (*n*)		Ethnicity (*n*)	
Female	312	Freshman	38	White	307
Male	139	Sophomore	120	Asian	78
		Junior	158	Africa-American	15
		Senior and beyond	135	Hispanic	4
				Native American	5
				Mixed	11

Results and Discussion

The descriptive statistics and inter-correlations of variables included in the path analysis are presented in the Table 19. The initial model included gender and student school year as manifested variables; however, there was no direct effect of the student gender and school year on student authentic research experiences, scientific literacy skills; therefore, these two variables were removed. The initial analyses of the model that excluded the variables of student gender and school year found that the direct effect of research courses and tools and techniques on scientific literacy were not significant so these paths were deleted in the final run. All remaining paths were significant in the final model (see Figure 11).

Table 19

Means, Standard Deviations, and Correlations of Variables included in Path Analysis

Variable	*M*	*SD*	1	2	3	4	5
1. Research Courses	.90	1.15					
2. Interest	4.15	.72	.146**				
3. Attitudes	3.93	.67	.194**	.543**			
4. Tools and techniques skills	3.98	.60	.237**	.393**	.696**		
5. Communication skills	3.89	.68	.234**	.345**	.581**	.819**	
6. Test of Science Literacy Skills	20.1	5.11	.086	.216**	.109*	.113*	.178**

*$p < .05$
**$p < .01$

The indices for the model were good indicating that the data fit well to the hypothesized model. The index of absolute fit, standardized root mean square error of approximation (RMSEA) was .058, smaller than the established threshold of .06. The comparative fit index (CFI) was .92, greater than the threshold of .90. The index of parsimonious fit, Chi Square /*df* was 2.70, greater than the threshold of .50. Since the model fit was good, we progressed to interpret the parameters in the measurement and

structural model. All parameters were significant in the reduced path model. The

relationship between the latent factor and its indicators was specified using the

measurement model, which showed that items clustered on each latent variable were

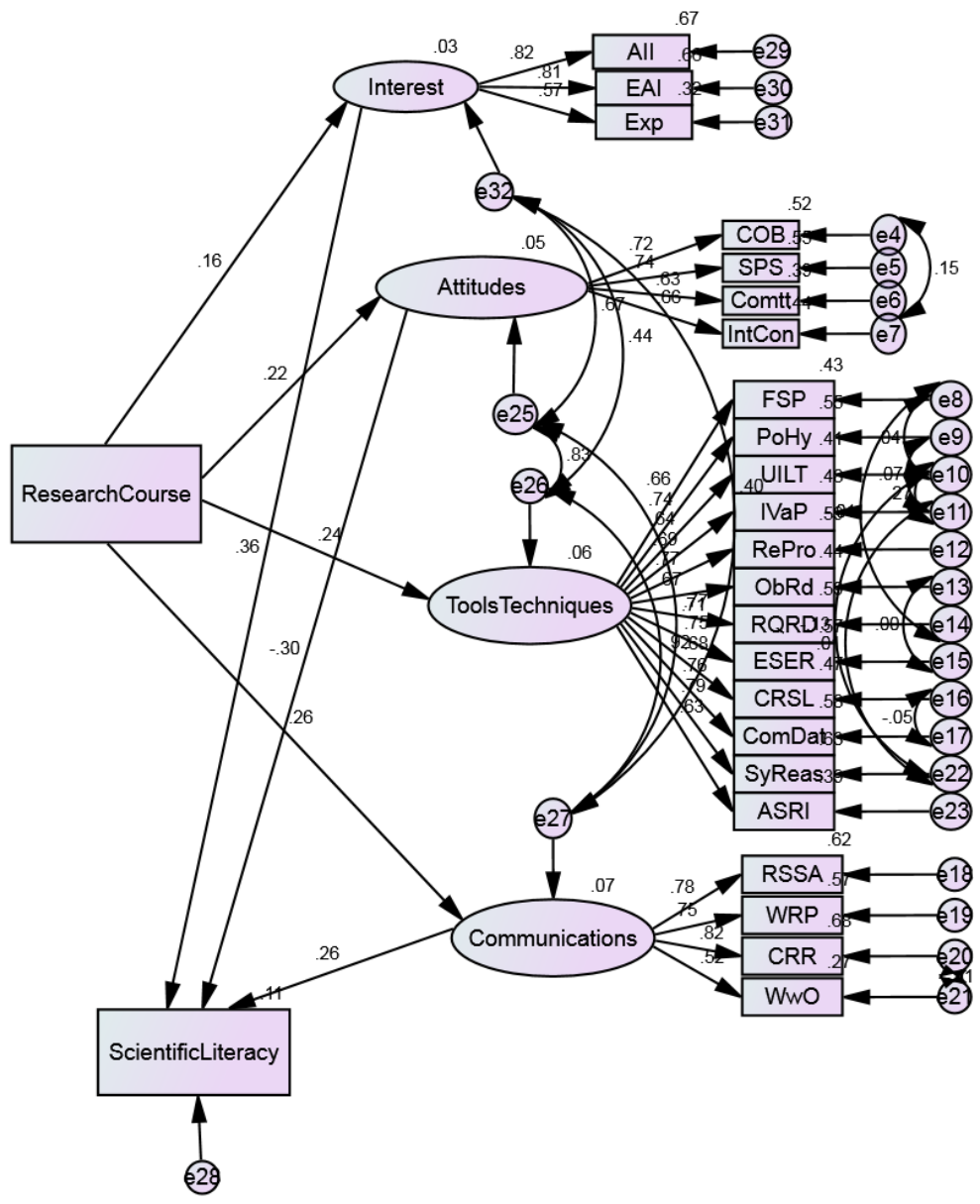significant, representing strong measures (see Figure 11).

Figure 11. Reduced Path Model of Student Scientific Literacy Level.

The effect of one manifest or latent variable on the other was interpreted as the structural relations in the model. All path coefficients were significant in the reduced path model. There were direct paths from interest, attitudes, and communication skills to the variable of scientific literacy skills, while attitudes had a negative direct relationship to the scientific literacy skills. The number of authentic research courses taken was not related to scientific literacy but had direct relationships to interest, attitudes, tools and techniques, and communication skills. The direct, indirect, and total effects on student scientific literacy level are presented in Table 20.

Table 20

Independent Variable	Interest			Attitudes			Tools and techniques skills			Communication skills			TOSLS scores		
	Drt	Ind	Ttl	Drt	Ind	Ttl	Drt	Ind	Ttl	Drt	Ind	Ttl	Drt	Ind	Ttl
Research Courses	.16*		.16*	.23*		.23*	.24*		.24*	.26*		.26*	.03	.06	.09
Interest													.36*		.36*
Attitudes													-.30*		-.30*
Communication skills													.25*		.26*

*Direct, Indirect, and Total Effects (*N = 450) on Science Literacy Level*

*p < .05. *Note.* Drt = Direct effect, Ind = Indirect effect, and Ttl= Total effect.

Results from the current investigation supported many of the predictive influences of participating in undergraduate authentic research courses on student science learning outcomes. Students who participated in more authentic research courses were predicted to show more interest in authentic scientific practice, have better attitudes toward doing science, show improved skills using tools and techniques, and evidence more effective communication. Student interest in authentic scientific inquiry had the highest direct relationship to scientific literacy with a path coefficient of .36. The path coefficient from

communication skills to scientific literacy was .25 suggesting a moderate influence. Unexpectedly, student self-reported skills regarding tool and technique use did not have a direct relationship to scientific literacy skills while attitudes had a negative relationship to scientific literacy.

The unexpected results regarding student attitudes and the lack of a reported relationship between tool and technique and scientific literacy may be due to measurement problems (Hackett, 1985) or to reduced exposure to scientific courses. A review of the data also found that some students who took over 15 science courses had not taken any authentic research courses whereas some students who took one or two authentic research courses had taken less than five science courses in total.

Therefore, it is possible that students engaged in authentic research experiences had less overall science training despite a relatively richer authentic research experience. This finding may impact scores on TOSLS since this test focuses on reading, interpreting and analyzing scientific data. In order to address the issue of the suspected covariate influence of the number of college science courses a student took, a multivariate analysis of covariance (MANCOVA) was conducted. The results showed that the number of college science courses a student took significantly predicted student attitudes, communication skills, tools and techniques, and scientific literacy ($p < .05$). Effect sizes (partial eta squared) were .02, .01, .01, .02 respectively. When the variable of number of college science courses was controlled, the student authentic research course experience still significantly predicted interest, attitudes, tools and techniques, and communication skills ($p < .05$) with effect sizes of .052, .054, .04, .03 respectively.

The purpose of this study was to investigate whether integrating authentic research in undergraduate science courses is achieving its educational goals and to provide evidence of educational gains for instructors hoping to adopt authentic research integrated curricula. Though a significant predictive influence of the number of authentic research experiences on student scientific literacy skills was not found, findings revealed a significant predictive influence of authentic research experiences on student interest in authentic research, attitudes toward participating in authentic scientific practice, tools and technique skills, and communication skills. Student interest in authentic research and communication skills are significant predictors of scientific literacy skills. The insights gained from this study will contribute to the lack of quantitative data in existence regarding the impact of course-based authentic inquiry experiences on student learning outcomes, especially student scientific literacy skills.

Chapter 7

Conclusion

This study investigated the impact of integrating authentic research projects in undergraduate science curricula on student learning outcomes. Scientific inquiry is the core of science education (NRC, 1996). Integrating original authentic research projects into undergraduate science curricula extends the opportunities of experiencing authentic scientific inquiry from to a few students selected and mentored by faculty to a large group of students enrolled in science courses (Auchincloss et al., 2014; Lee & Songer, 2003; Linn, Palmer, Baranger, Gerard, & Stone, 2015). Few empirical studies that assess the impact of course-based authentic inquiry model exist. Validity related problems are big concern in the assessment of undergraduate research experiences (Auchincloss et al., 2014; Wei & Woodin, 2011).

Scientific literacy skills, which is an important goal of science education (NRC, 1996), was not studied as an outcome of integrating authentic research projects in undergraduate science curriculum. This study is the first effort that used validated instruments to investigate the predictive influence of student course-based authentic research experiences on scientific literacy skill. Student interest in science and authentic scientific inquiry competency were mediator variables. Authentic scientific inquiry competency in this study refers to three sub-categories: attitudes toward doing science, skills of using tools and techniques, and communication skills. This chapter presents the summary of the findings from previous chapters, as well as discusses the limitations of the current study and recommendations for future research.

Summary of Findings

This study used path analysis, a form of Structural Equation Modelling, to investigate the predictive influence of course-based authentic inquiry experiences on student learning outcomes including student scientific literacy skills, interest in doing science, and scientific inquiry competency. The number of authentic research courses a student took was the indicator of the level of student authentic research experiences. In general, there was no statistically significant effect of student authentic research experience on student scientific literacy skills. There were, however, significant direct effects on student interest in doing science and scientific inquiry competency.

Student interest in authentic scientific inquiry had the highest direct effect on scientific literacy skills with a path coefficient of .36. The significant effect of student communication skills on scientific literacy skills was moderate with a path coefficient of .25. These findings add inferential evidences to the positive benefits of undergraduate research experiences on student learning that are described in previous studies (e.g.,; Bergwerff & Warners, 2007; Bernard, 2011; Bussey et al., 2015; Campbell et al., 2012; Hartmann et al., 2013; Hunter et al., 2006; Jaarsma et al., 2009; Thiry, Weston, Laursen, & Hunter, 2012; Luckie et al., 2012; Miller et al., 2013; Nadelson et al., 2010; Naug et al., 2012; Nugent et al., 2008).

Nevertheless, due to the lack of inferential studies that examined the effect size of the influence of undergraduate research experiences, especially course-based authentic research experiences, it is hard to compare the effect sizes as a result of the path analysis conducted in this study. Meanwhile, since there is no literature that examined the impact of undergraduate research experiences on student scientific literacy skills, the results of

this study regarding the impact of course-based authentic research on student scientific literacy skills may provide references to future study for comparison and discussion.

Evidence from this study are expected to encourage instructors who seek to adopt and implement authentic scientific inquiry-based curricula as a way to improve undergraduate science education. In addition, the findings from this study may provide science educators who are interested in reforming science education with some insights of the implementation and values of an instructional model that integrates authentic research into undergraduate science curriculum.

In addition to investigating the predictive influence of integrating authentic research projects in undergraduate science curricula, this study developed and validated a new instrument entitled Student Science Learning Gains (SSLG) survey for specifically assessing the influence of integrating authentic research projects in undergraduate science curriculum. The lack of inclusive evidence regarding the impact of integrating authentic research projects in undergraduate science is a result of two main problems. One is the lack of inferential studies and the other one is the validity related issues in the assessment of undergraduate research experiences.

Most quantitative studies provide little information about the reliability and validity of the instruments used for data collection. This study used exploratory factor analysis and confirmatory factor analysis to validate SSLG survey. In particular, the construct validity, which is rarely studied in instrument development, was tested and reached in this study. The effort of developing and validating of SSLG will contribute to the survey research in higher education. This SSLG instrument lends practical significance for program assessment regarding authentic scientific inquiry-based

curriculum and instruction. The validated instrument is ready to be used to assess the impact of authentic research integrated into undergraduate science curriculum, a goal that is advocated and funded by the National Science Foundation since the 1980s. It also helps distinguish student gains from authentic and closed investigations within the classroom as requested by scholars (i.e., Hume, 2009; Schwartz, Lederman, & Crawford, 2004).

Another contribution of this study to existing literature regarding the assessment of undergraduate research experiences is that this study conducted a systematic literature review of the impact of undergraduate research. The systematic literature review examined current studies in terms of study trends, quality of the study, undergraduate research delivery forms, assessment design, the authenticity of the inquiry projects and research experiences, the evidence of the impact of undergraduate research experiences, and the validation of instruments used in these assessment studies.

The findings of the systematic literature review indicate that, course-based undergraduate research experiences have receiving increasing interest and popularity in the past several years. Studies revealed that undergraduate research experiences have positive impact on student learning outcomes in a variety of ways. However, inferential study, especially controlled study is rare. Validity related problems are concerned in studies. In addition, the authenticity of student inquiry experiences is ignored in most studies. Applied situated learning theory and cognitive apprenticeship model, authentic learning context is the core to effective scientific inquiry instructional design and implementation. These findings suggest the need of control studies that can provide

evidence that student benefits from participating undergraduate research experiences are significant higher than traditional science education.

Limitations of the Study

Previous studies that assessed undergraduate research experiences showed diverse limitations including lack of a clear definition of inquiry, ignoring the feature of authenticity of inquiry experiences, validity related issues of the study design and measures, lack of study that examined the impact of undergraduate research on student scientific literacy skill, and lack of assessments that generated conclusive and inferential results. This dissertation aimed at reducing several of those limitations by proving a clear framework for what authentic inquiry means in this study from the perspective of the authentic context of the inquiry activity, and the role of students in authentic inquiry learning; developing and validating a new instrument that is specifically used to assess student learning gains from participating authentic research that is integrated in undergraduate science curricula; applying path analysis to examine the predictive influence of student authentic undergraduate research experiences on student scientific literacy skills, interest in doing science, and scientific inquiry competency.

Nevertheless, this study has a few limitations. Even though the confirmatory factor analysis yielded good estimates of the construct validity of SSLG survey, certain limitations in this study should be considered when others attempt to apply the SSLG instrument in authentic scientific inquiry related program assessment. Achieving validity and reliability is the first step in instrument development. The SSLG survey is a student self-reported instrument; therefore, the issue of subjectivity has to be taken into account when it is used. Triangulation via mixed methods, such as mixing the use of survey data

with other quantitative and qualitative data, is an approach to continued validation (Jick, 1979). Faculty involved in the SSLG development is from chemistry and biology related domains so that this instrument may better function in biology and chemistry related authentic research courses. Additional work in other science domains is suggested. The data generated from the SSLG survey could provide valuable information with regard to student science learning gains along with other assessment resources, such as data collected from survey or interviews with instructors, student interviews, and student craftworks.

Another limitation of this study is the sample selection and the study design. The study used convenient sampling selection, which suffers from a number of biases. The convenience sample can lead to the under-representation or over-representation of particular groups within the sample. In addition, since the sampling frame is unknown, and the sample is not chosen at random, it is uncertain that the sample would be representative of the population being studied. This weakness limits the ability of this study to generalization from this sample to the population of students who are involved in undergraduate research experiences.

This study used path analysis, a form of Structural Equation Model (SEM) to investigate the relationship between student course-based authentic research experiences on student scientific literacy skills, interest in doing science, and scientific inquiry competency. Though path analysis is a technique to evaluate causal hypotheses, it cannot establish the direction of causality. In addition, the results of path analysis showed moderate effect of course-based undergraduate research experiences on student interest in doing science, attitudes toward doing scientific inquiry skills, skills of using tools and

techniques, and communication skills, it cannot tell that students had significant higher gains from participating in authentic research than the gains they would have from traditional science instruction.

Another limitation of this study is that, the impact of authentic research experiences on content knowledge comprehension is not included due to the lack of validated measures of student content knowledge in a few different science courses.

Recommendations for Future Research

The results obtained from this dissertation can serve as a stimulus for future research on the impact of course-based undergraduate research experiences. First, this study did not find a significant direct effect of student course-based authentic research experiences on student literacy skills. However, it does not mean involving students in authentic research would not improve student literacy skills. A possible reason might be that, the Test of Student Scientific Literacy (TOSSL) requires a broad scientific knowledge and information, but freshman or sophomore students who have taken a few authentic research courses have not build enough knowledge to conduct the TOSSL well. This finding suggests more research in the future to investigate the impact of undergraduate authentic research experiences on scientific literacy skills with the use of different methods or instruments. In addition, this study revealed a moderate effect of course-based undergraduate research experiences on student learning outcomes, however, it did not compare student learning gains from research-based science course and from the traditional science course. The control study is suggested in the future research. Assessing other benefits of course-based authentic research experiences such as content knowledge learning and application, student involvement in the culture of science,

identification of researcher, and the development of view of the nature of science, using validated instruments are recommended in future research.

In conclusion, this study attempted to bridge some research gaps in the field of undergraduate research experiences. The assessment of student learning outcomes from participating in authentic research courses has practical significance for providing insights and data-driven evidence for decision making in educational reform. In this study, integrating authentic research projects is suggested as an effective, economic, and realistic approach to engaging larger student population in authentic inquiry. The assessment of the curriculum model that integrates authentic research projects using validated instruments revealed moderate but significant association between students' authentic research experiences on scientific literacy skills, interest, attitudes toward doing science. The assessment data and findings generated from this study are expected to help instructors seeking to expand portions of their traditional science curriculum to include authentic research. In doing so, they will enhance student learning and stimulate engagement. Immersing students in the collaborative process of authentic scientific inquiry, from development to publication, prepares students for future careers, stimulates instructor engagement, and provides meaningful novel data to the larger scientific community.

References

Adedokun, O. A., & Burgess, W. D. (2011). Uncovering students' preconceptions of undergraduate research experiences. *Journal of STEM Education, 12*(5), 12-21.

Adedokun, O. A., Zhang, D., Parker, L. C., Bessenbacher, A., Childress , A., & Burgess, W. D. (2012). Understanding how undergraduate research experiences influence student aspirations for research careers and graduate education. *Journal of College Science Teaching, 42*(1), 82-90.

Ahn, A. C., Tewari, M., Chi-Sang, P., & Phillips, R. S. (2006). The limits of reductionism in medicine: Could systems biology offer an alternative?. *Plos Medicine*, *3*(5), e208-0713.

Aikenhead, G. S., & Ryan, A. G. 1992. The development of a new instrument: Views on Science-Technology-Society (VOSTS). *Science Education, 76*(5). 477-491.

American Association for the Advancement of Science. (1989). *Project 2061: Science for all Americans.* New York, NY: Oxford University Press.

American Association for the Advancement of Science. (1993). *Benchmarks for science literacy.* New York, NY: Oxford University Press.

Anderson, R. D. (2002). Reforming science teaching: What research says about inquiry. *Journal of Science Teacher Education, 13*(1), 1-12.

Atkin, J. M., & Black, P. (2003). *Inside science education reform: A history of curricular and policy change.* New York, NY: Teachers College Press.

Auchincloss, L. C., Laursen, S. L., Branchaw, J. L., Eagan, K., Graham, M., Hanauer, D. I., . . . Dolan, E. L. (2014). Assessment of course-based undergraduate research experiences: A meeting report. *CBE-Life Sciences Education, 13*(1), 29-40.

Barab, S. A., Squire, K. D., & Dueber, W. (2000). A coevolutionary model for supporting the emergence of authenticity. *Educational Technology Research and Development, 48*(2), 37-62.

Barker, L. (2009). Student and faculty perceptions of undergraduate research experiences in computing. *ACM Transactions on Computing Education, 9*(1), 1-28.

Bearman, M., Smith, C. D., Carbone, A., Slade, S., Baik, C., Hughes-Warrnington, M., & Neumann, D. L. (2012). Systematic review methodology in higher education. *Higher Education Research & Development, 31*(5), 624-620.

Behar-Horenstein, L. S., & Johnson, M. L. (2010). Enticing students to enter into undergraduate research: The instrumentality of an undergraduate course. *Journal of College Science Teaching, 39*(3), 62-70.

Bell, R. L., Blair, L. M., Grawford, B. A., & Lederman, N. G. (2003). Just do it? Impact of a science apprenticeship program on high school students' understandings of the nature of science and scientific inquiry. *Journal of Research in Science Teaching, 40*(5) 487-509.

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

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

Bergwerff, K., & Warners, D. (2007). Multiple objectives achieved with a germination experiment in a science education biology class. *American Biology Teacher, 69*(9), 552-556.

Bernard, W. (2011). What students really think about doing research. *Science Teacher, 78*(8), 52-54.

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

Bonaccorsi, A. (2010). New forms of complementarity in science. *Minerva: A Review of Science, Learning & Policy*, *48*(4), 355-387.

Briggs, M., Long, G., & Owens, K. (2011). Qualitative assessment of inquiry-based teaching methods. *Journal of Chemical Education* (88), 1034-1040.

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18*(1), 32-42.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen, & J. S. Long (Eds.), *Testing structural equation models* (136-162), Newbury Park, CA: Sage.

Bruner, J. (1960). *The process of education.* Cambridge, MA: Harvard University Press.

Buck, L. B., Bretz, S. L., & Towns, M. H. (2008). Characterizing the level of inquiry in the undergraduate laborary. *Journal of College Science Teaching, 38*(1), 52-58.

Bussey, K. A., Cavalier, A. R., Connell, J. R., Mraz, M. E., Holderread, A. S., Oshin, K. D., & Pintauer, T. (2015). Synthesis and characterization of copper complexes with a tridentate nitrogen-donor ligand: An integrated research experiment for undergraduate students. *Journal of Chemistry Education, 92*(12), 2140-2145.

Cakir, M. (2011). Enhancing mendelian genetics concepts using a guided computer-mediated inquiry. *Journal of Baltic Science Education, 10*(3), 156-167.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*(2), 81-105.

Campbell, T., Abd-Hamid, N., & Chapman, H. (2010). Development of instruments to assess teacher and student perceptions of inquiry experiences in science classrooms. *Journal of Science Teacher Education*, *21*(1), 13-30.

Campbell, T., Wolf, P. G., Der, J. P., Packenham, E., & Abd-Hamid, N. (2012). Scientific inquiry in the genetics laboratory: Biologists and university science teacher educators collaborating to increase engagements in science processes. *Journal of College Science Teaching, 41*(3), 82-89.

Canaria, J. A., Schoffstall, A. M., Weiss, D. J., Henry, R. M., & Braun-Sand, S. B. (2012). A model for an introductory undergraduate research experience. *Journal of Chemical Education, 89* (11), 1371-1378.

Capps, D. K., Crawford, B. A., & Constas, M. A. (2012). A review of empirical literature on inquiry professional development: Alignment with best practices and a critique of the findings. *Journal of Science Teacher Education*, *23*(3), 291-318.

Chang, H., Chen, C., Guo, G., Cheng, Y., Lin, C., & Jen, T. (2011). The development of a competence scale for learning science: Inquiry and communication. *International Journal of Science and Mathematics Education,* (5), 1213-1233.

Chinn, C. A., & Malhotra, B. A. (2002). Epistemologically authentic inquiry in schools: A theoretical framework for evaluating inquiry tasks, *86*, 175-218.

Choi, J. & Hannafin, M.J. (1995). Situated cognition and learning environments: roles, structures, and implications for design. *Educational Technology Research and Development, 43 (2)*, 53—69.

Chung, H.-M., & Behan, K. J. (2010). Peer sharing facilitates the effect of inquiry-based projects on science learning. *American Biology Teacher, 72*(1), 24-29.

Collins, A. (1988). *Cognitive apprenticeship and instructional technology (Technical Report 6899).* Cambridge, MA: BBN Labs, Inc.

Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship:Teaching the craft of reading, writing, and mathematics. In L. B. Resnick, *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 453-494). Hillsdale, NJ: Lawrence Erlbaum Associates.

Corwin, L. A., Graham, M. J., & Dolan, E. L. (2015). Modeling course-based undergraduate research experiences: An agenda for future research and evaluation. *CBE-Life Sciences Education, 14*(1), 1-13.

Costello, A. B., & Osborne, J. W. (2005). Exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research, and Evaluation, 10*(7), 1-9.

Coverdale, G.A.(2002). Tracking Development: Pursuing Scientific Inquiry on the Lytle Creek. *Journal of College Science Teaching, 32*(3), 1880193.

Crowe, M., & Brakke, D. (2008). Assessing the impact of undergraduate research experiences on students: An overview of current literature. *Council on Undergraduate Research Quarterly, 28*(4), 43-50.

Culp, B., & Urtel, M. (2013). Demonstrated successful undergraduate research experiences across the dicsiplines - the physical education teacher education perspective. *Journal of Physical Education, Recreation & Dance, 84*(9), 24-27.

Davidson, Z. E., & Palermo, C. (2015). Developing research competence in
undergraduate students through hands on learning. *Journal of Biomedical
Education, 2015*, 1-9.

Davies, D., Jindal-Snape, D., Collier, C., Digby, R., Hay, P., & Howe, A. (2013).
Creative learning environments in education—A systematic. *Thinking Skills and
Creativity, 8*(1), 80-91.

Dennen, V. P. (2004). Cognitive apprenticeship in educational practice: Research on
scaffolding, modeling, mentoring, and coaching as instructional strategies. In D.
H. Jonassen, *Handbook of research on educational communication and
technology* (2nd ed., pp. 813-828). Mahwah, NJ: US: Lawrence Erlbaum
Association Publishers.

Dewey, J. (1933). *How we think* (2nd ed.). Boston, MA: D.C.Heath.

Dillner, D. K., Ferrante, R. F., Fitzgerald, J. P., & Schroeder, M. J. (2011). Integrated
laboratories: Laying the foundation for undergraduate. *Journal of Chemistry
Education, 88*(12), 1623-1629.

Dodick, J. & Orion, N (2003). Measuring student understanding of geological time.
Science Education, 87(5). 708-731.

Duschl, R. A., Schweingruber, H. A. & Shouse, A. W. (Eds.) (2007). *Taking science to
school: Learning and teaching science in grades K–8.* Washington, DC: National
Academies Press.

Edelson, D. C. (1998). Realising authentic science learning through the adaptation of
scientific practice. In B. J. Fraser, & K. G. Tobin (Eds.), *Internation Handbook of
Science Education* (pp. 317-331). Great Britain: Kluwer Academic Publishers.

Ellis-Monaghan, J., & Pangborn, G. (20142013). An example of practical organization for undergraduate research experiences. *PRIMUS: Problems, Resources, and Issues in Mathematics Undergraduate Studies, 23*(9), 805-814.

Enkenberg, J. (2001). Instructional design and emerging models in higher education. *Computers in Human Behavior, 17*, 495–506.

Fang, F. C., & Casadevall, A. (2011). Reductionistic and holistic science. *Infection and Immunity, 79*(4), 1401– 1404.

Farrell, A. M. (2010). Insufficient discriminant validity: A comment on Bove, Pervan, Beatty and Shiu (2009). *Journal of Business Research, (63)*3, 324-327.

Fensham, P. J. (2002). Time to change drivers for scientific literacy. *Canadian Journal of Science, Mathematics and Technology Education, 2*(1), 9-24.

Field, A. P. (2009). *Discovering statistics using SPSS* (3rd ed.). London, UK: Sage.

Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment, 7*(3), 286-299.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18*(1), 39-50.

Gilmore, J., Vieyra, M., Timmerman, B., Feldon, D., & Maher, M. (2015). The relationship between undergraduate research participation and subsequent research performance of early career STEM graduate students. *Journal of Higher Education, 86*(6), 834-863.

Godek, Y., Kaya, V. H., & Polat. D (2015). The relationship between Turkish primary

    school students' scientific literacy levels and scientific process skills. *Cypriot*

    Journal of Educational Sciences, 10(1), 3.

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.

Gormally, C., Brickman, P., Hallar, B., & Armstrong, N. (2011). Lessons learned about

    implementing an inquiry-based curriculum in a college biology laboratory

    classroom. *Journal of College Science Teaching, 40*(3), 45-51.

Greeno, J. (1997). On claims that answer the wrong questions, *Educational Researcher.*

    26(1), 5-17.

Griffard, P. B., & Golkowska, K. (2013). Early exposure to research: Outcomes of the

    ASTER certification program. *Journal of College Science Teaching, 43*(1), 37-42.

Hackett, G. (1985). Role of mathematics self-efficacy in the choice of math-related

    majors of college women and men: A path analysis. *Journal of Counseling*

    Psychology, 32(1), 47-56.

Hair, J .F. Jr., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate data*

    analysis (5th ed.). Upper Saddle River, NJ: Prentice Hall.

Hanauer, D. I., & Hatfull, G. (2015). Measuring networking as an outcome variable in

    undergraduate research experiences. *CBE-Life Science Education, 14*(4), 1-10.

Hanauer, D. I., Frederick, J., Fotinakes, B., & Strobel, S. A. (2012). Linguistic analysis of

    project ownership for undergraduate research experiences. *CBE-Life Science*

    Education, 11(4), 378-385.

Hanauer, D. I., Hatfull, G., & Jacobs-Sera, D. (2009). *Active assessment: Assessing scientific inquiry*. New York, NY: Springer-Verlag.

Harlen, W. (1999). Purposes and procedures for assessing science process skills. *Assessment in Education, 6*(1), 129-144.

Harsh, J. A., Maltese, A. V., & Tai, R. H. (2011). Undergraduate research experiences from a longitudinal perspective. *Journal of College Science Teaching, 4*(1), 84-91.

Harsh, J. A., Maltese, A. V., & Tai, R. H. (2012). A perspective of gender differences in chemistry and physics undergraduate research experiences. *Journal of Chemistry Education, 89*(11), 1364-1370.

Hartmann, A. C. (2013). Education: Science literacy benefits all. *Nature*, *498*(7453), 171.

Hartmann, Q. J., Widner, S. C., & Carrick, C. (2013). Strong faculty relationships and academic motivation as potential outcomes of undergraduate research. *North American Journal of Psychology, 15*(1), 215-234.

Hernandez, R. (2010). A short form of the Subjective Well-Being Scale for Filipinos. *Educational Measurement and Evaluation Review, 1*, 105-115.

Herrington, J., & Oliver, R. (2000). An instructional design framework for authentic learning environments. *Educational Technology Research and Development, 48*(3), 23-48.

Hieronymi, A. (2013). Understanding systems science: A visual and integrative approach. *Systems Research & Behavioral Science, 30*(5), 580-596.

Hodson, D. (1992). Assessment of practical work: Some considerations in philosophy of science. *Science & Education, 1*(2), 115-144.

Hodson, D. (1996). Laboratory work as scientific method: three decades of confusion and distortion. *Journal of Curriculum Studies, 28*(2), 115-135

Hofstein, A., & Lunetta, V. N. (20042003). The laboratory in science education: Foundations for the twenty - first century. *Science Education, 88*(1), 28-54.

Hotho, J., Saka-Helmhout, A., & Becker-Ritterspach, F. (2014). Bringing context and strucure back into situated learning. *Management Learning, 45*(1), 57-80.

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

Hume, A. (2009). Authentic scientific inquiry and school science. *Teaching Science: The Journal of the Australian Science Teachers Association, 55*(2), 35-41.

Hunter, A.-B., Laursen , S. L., & Seymour, E. (20072006). Becoming a scientist: The role of undergraduate research in students' cognitive, personal, and professional development. *Science Education*, 36-74.

Iimoto, D. S., & Frederick, K. A. (2011). Incorporating student-designed research projects in the chemistry curriculum. *Journal of Chemical Education, 88*(8), 1069-1073.

Ing, M., Fung, W. W., & Kisailus, D. (2013). The influence of materials science and engineering undergraduate research experiences on public communication skills. *Journal of STEM Education: Innovations and Research, 14*(2), 16-20.

Jaarsma, D.A, Muijtjens, A.M., Dolmans, D.H., Schuurmans, E.V., Beukelen, P.V., & Scherpbier, J.A. (2009). Undergraduate research internships: Veterinary students'

experiences and the relation with internship quality. *Medical Teacher, 31* (5), e178-e184.

Jansen, D., Jadack, R. A., Ayoola, A. B., Doornbos, M. M., Dunn, S. L., Moch, S. D., . . . Wegner, G. D. (2015). Embedding research in undergraduate learning opportunities. *Western Journal of Nursing Research, 37*(10), 1340-1358.

Jick, T. D. (1979). Mixing qualitative and quantitative methods: Triangulation in action. *Administrative Science Quarterly, 24*(4), 602-611.

Jones, M. T., Barlow, A. E., & Villarejo, M. (2010). Importance of undergraduate research for minority persistence. *The Journal of Higher Education, 81*(1), 82-115.

Joreskog, K. G. & Sorbom, D. (1984). *LISREL-VI user's guide* (3rd ed.). Mooresville, IN: Scientific Software.

Kardash, C. M., & Edwards , O. V. (2012). Thinking and Behaving like scientists: Perceptions of undergraduate science interns and their faculty mentors. *Instructional Science: An International Journal of the Learning Sciences, 40*(6), 875-899.

Kendricks, K., & Arment, A. (2011). Adopting a K-12 family model with undergraduate research to enhance STEM persistence and achievement in underrepresented minority students. *Journal of College Science Teaching, 41*(2), 22-27.

Kim, K. A., Fann, A. J., & Misa-Escalante, K. O. (2011). Engaging women in computer science and engineering: promising practices for promoting gender equity in undergraduate research experiences. *ACM Transactions on Computing Education (TOCE), 11*(2), 8:1-8:13.

Kuhn, D. (1997). Constraints or guideposts? Developmental psychology and science education. *Review of Educational Research, 67*(1), 141-150.

Kuhn, T. (1996). *The structure of scientific revolutions* (3rd ed.). Chicago, IL: University of Chicago Press.

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

Lee, H.-S., & Songer, N. B. (2003). Making authentic science accessible to students. *International Journal of Science Education, 25*(8), 923-948.

Lim, J., Reiser, R. A., & Olina, Z. (2009). The effects of part-task and whole-task instructional approaches on acquisition and transfer of a complex cognitive skill. *Educational Technology and Development, 57*(1), 61-77.

Linn, M. C., Palmer, E., Baranger, A., Gerard, E., & Stone, E. (2015). Undergraduate research experiences: Impacts and opportunities. *Science, 347*(6222), 627–633.

Lopatto, D. (2004). Survey of Undergraduate Research Experiences (SURE): First findings. *Cell Biology Education, 3*(4), 270-277.

Lopatto, D. (2007). Undergraduate research experiences support science career decisions and active learning. *CBE-Life Science Education, 6*(4), 297–306.

Lopatto, D. (2010). Undergraduate research as a high-impact student experience. *Association of American Colleges and Universities, 12*(2), 27-30.

Lopatto, D., & Tobias, S (2010). *Science in solution: The impact of undergraduate research on student learning*, Washington, DC: Council on Undergraduate Research.

Luckie , D. B., Aubry, J. R., Marengo , B. J., Rivkin, A. M., Foos, L. A., & Maleszewski , J. J. (2012). Less teaching, more learning: 10-yr study supports increasing student learning through less coverage and more inquiry. *American Journal of Physiology - Advances in Physiology Education, 36*(4), 325-335.

Lustick, D. (2009). The failure of inquiry: Preparing science teachers with an authentic investigation. *Journal of Science Teacher Education, 20*(6), 583-604.

Marsh, H. W., & Hocevar, D. (1985). Application of confirmatory factor analysis to the study of self-concept: First-and higher order factor models and their invariance across groups. *Psychological Bulletin, 97*(3), 562-582

Miller, C. W., Hamel, J., Holmes, K. D., Helmey-Hartman, W. L., & Lopatto, D. (2013). Extending your research team: Learning benefits when a laboratory partners with a classroom. *BioScience, 63*(9), 754-762.

Miller, H. R., McNeal, K. S., & Herbert, B. E. (2010). Inquiry in the physical geology classroom: Supporting students' conceptual model development. *Journal of Geography in Higher Education, 34*(4), 596-615.

Nadelson, L. S., Walters, L., & Waterman, J. (2010). Course-integrated undergraduate research experiences structured at different levels of inquiry. *Journal of STEM Education, 11*(1&2), 27-44.

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

National Science Teachers Association (NSTA). (2007). *NSTA position statement. The Integral role of laboratory investigations in science instruction.* Retrieved from: http://www.nsta.org/about/positions/laboratory.aspx

Naug, H. L., Colson, N. J., & Donner, D. G. (2012). The research encounter: An innovative course inclusion that facilitates student engagement. *Innovative Higher Education, 37*(4), 335-345.

Niewolny, K. L., & Wilson, A. L. (2009). What happened to the promise: A critical (re)orientation of two sociocultural learning traditions. *Adult Education Quarterly, 60*(1), 26-45.

Nugent, G., Kunz , G., Levy, R., Harwood, D., & Carlson, D. (2008). The impact of a field-based, inquiry-focused model of instruction on preservice teachers' science learning and attitudes. *Electronic Journal of Science Education, 12*(2), 94-111.

Nugent, G., Toland, M. D., Levy, R., Kunz, G., Harwood, D., Green, D., & Kitts, K. (2012). The impact of an inquiry-based geoscience field course on pre-service teachers. *Journal of Science Teacher Education, 23*(5), 503-529.

Nunnally, J. C., & Bernstein, I. H. (1994). Psychometric theory (3rd ed.). New York, NY: McGraw-Hill.

Pacifici, L. B., & Thomson, N. (2011). What do they expect? A comparison of student expectations and outcomes of undergraduate research experiences. *Journal of College Science Teaching, 41*(1), 54-60.

Pender, M., Marcotte, D. E., Domingo, M. R., & Maton, K. I. (2011). The STEM pipeline: The role of summer research experience in minority students' Ph.D. aspirations. *Education Policy Analysis Archives, 18*(30), 1-36.

Petticrew, M., & Roberts, H. (2006). *Systematic reviews in the social science: A practical guide*. Malden, MA: Blackwell.

Popper, K. (1959). Conjectures and Refutations, London.

Powell, N. L., & Harmon, B. B. (2014). Developing scientists: A multiyear research experience at a two-year college. *Journal of College Science Teaching, 44*(2), 11-17.

Quardokus, K., Lasher-Trapp, S., & Riggs, E.M. (2012). A Successful Introduction of Authentic Research Early in an Undergraduate Atmospheric Science Program. *Bulletin of the American Meteorological Society, 93*(11), 1641-1649.

Rapoport, A. (1986). *General system theory*: *Essential concepts and applications.* Cybernetics and systems series. Tunbridge Wells, UK: Abacus Press.

Reiser, B. J., Radinsky, J., Edelson, D. C., Gomez, L. M., & Marshall, S. (2001). Developing reflective inquiry practices: A case study of software, the teacher, and students. In K. Crowley, C. D. Schunn, & T. Okada (Eds.), (Tran.), *Designing for science: Implications from everyday, classroom, and professional settings* (247-286). Mahwah, NJ: Lawrence Erlbaum Associates.

Rudolph, J. L. (2000). Reconsidering the 'nature of science' as a curriculum component. *Journal of Curriculum Studies, 32*(3), 403 – 419.

Russell, J. E., D'Costa, A. R., Runck, C., Barnes, D. W., Barrera, A. L., Hurst-Kennedy, J., . . . Schlueter, M. (2015). Bridging the undergraduate curriculum using an integrated course-embedded undergraduate research experience (ICURE). *CBE—Life Sciences Education, 14*, 1-10.

Russell, S. H., Hancock, M. P., & McCullough, J. (2007). Benefits of undergraduate research experiences. *Science, 316*(5824), 548-549.

Sadler, T. D., & McKkinney, L. (2010). Scientific research for undergraduate students: A review of the literature. *Journal of College Sceince Teaching, 39*(5), 43-49.

Sadler, T. D., Burgin, S., McKinney, L., & Ponjuan, L. (2010). Learning science through research apprenticeships: A critical review of the literature. *Journal of Research in Science Teaching, 47*(3), 235-256.

Schussler, E. E., Bautista, N. U., Link-Pérez, M. A., Solomon, N. G., & Steinly, B. A. (20142013). Instruction matters for nature of science understanding in college biology laboratories. *BioScience, 13*(3), 387-396.

Schwartz, R. S., Lederman, N. G., & Crawford, B. A. (2004). Developing views of nature of science in an authentic context: An explicit approach to bridging the gap between nature of science and scientific inquiry. *Science Education, 88*(4), 610-645.

Schwarzer, R., & Jerusalem, M. (1995). Generalized Self-Efficacy scale. In J. Weinman, S. Wright, & M. Johnston, *Measures in health psychology: A user's portfolio. Causal and control beliefs* (pp. 35-37). Windsor, UK: NFER-NELSON.

Seymour, E., Hunter, A.-B., Laursen, S. L., & DeAntoni, T. (2004). Establishing the benefits of research experiences for undergraduates in the sciences: First findings from a three-year study. *Science Education, 88*(4), 493-534.

Seymour, E., Wiese, D., Hunter, A. & Daffinrud, S. M. (2000). *Creating a better mousetrap: On-line student assessment of their learning gains.* Paper presentation at the National Meeting of the American Chemical Society, San Francisco, CA. Retrieved from: http://www.salgsite.org/docs/SALGPaperPresentationAtACS.pdf

Shanle, E. K., Tsun, I. K., & Strahl, B. D. (2016). A course-based undergraduate research experience investigating 300 bromodomain mutations. *Biochem Mol Biol Educ., 44*(1), 68-74.

Shields, G. C., Hewitt, G. J., & North, L. (2010). Using pre-college research to promote student success and increase the number of science majors. *Council on Undergraduate Research Quarterly, 31*(1), 43-47.

Shih, B.-Y., Chen, T.-H., Wang, S.-M., & Chen, C.-Y. (2013). The exploration of applying LEGO NXT in the situates science and technology learning. *Journal of Baltic Science Education, 12*(1), 73-91.

Somerville, C. R., Bauer, S, Brininstool, G., Facette, M., Hamann, T., Milne, J., … Youngs, H. (2004). Towards a systems approach to understanding plant cell walls. *Science, 306*, 2206-2211. Retrieved from: http://works.bepress.com/ted_k_raab/5

Song, Y., & Schwenz, R. (2013). An inquiry-based approach to teaching the spherical earth model to preservice teachers using the global positioning system. *Journal of College Science Teaching, 42*(4), 50-58.

Stevens, J. (2009). *Applied multivariate statistics for the social sciences* (5th ed.). New York, NY: Routledge.

Strobel, J., Wang, J., Weber, N. R., & Dyehouse, M. (2013). The role of authenticity in design-based learning environments: The case of engineering education. *Computers & Education, 64*, 143-152.

Tanaka, J. S., & Huba, G. J. (1985). A fit index for covariance structure models under arbitrary GLS estimation. *British Journal of Mathematical and Psychology, 38*. 197-201.

Taraban, R., & Logue, E. (2012). Academic factors that affect undergraduate research experiences. *Journal of Educational Psychology, 104*(2), 499-514.

Taraban, R., & Logue, E. (2012). Academic factors that affect undergraduate research experiences. *American Psychological Association, 104*(2), 499-514.

Thiry, H., Weston, T. J., Laursen, S. L., & Hunter, A.-B. (2012). The benefits of multi-year research experiences: differences in novice and experienced students' reported gains from undergraduate research. *CBE-Life Science Education, 11*(3), 260-272.

Timmerman, B. E., Strickland, D. C., Johnson, R. L., & Payne, J. R. (2011). Development of a "universal" rubric for assessing undergraduates' scientific reasoning skills using scientific writing. *Assessment & Evaluation in Higher Education, 36*(5), 509-547.

Tytler, R. (2007). Re-imagining science education: Engaging students in science for Australia's future. Retrieved from http://research.acer.edu.au/aer/3/

Urias, D., Gallagher, P., & Wartman, J. (2012). Critical features and value in assessing a research experience for undergraduates: The case of engineering cities. *Journal of STEM Education: Innovations & Research*, *13*(1), 30-42.

Utley, B. L. (2006). Effects of situated learning on knowledge gain of instructional strategies by students in graduate level course. *Teacher Education and Special Education: The Journal of the Teacher Education Division of the Council for Exceptional Children, 29*(1), 69-82.

Vieyra, M., Carlson, A., Leaver, E., & Timmerman, B. (2013). Undergraduate research: I am not sure what it is, but I don't have time to do it anyway. *Council on Undergraduate Research Quarterly, 33*(3), 27-34.

Vieyra, M., Gilmore, J., & Timmerman, B. (2011). Requiring research may improve retention in STEM fields for underrepresented women. *Council on Undergraduate Research Quarterly, 32*(1), 13-19.

Vygotsky, L. S. (1978). *Mind in Society.* Cambridge, MA: Harvard University Press.

Wagner, J. R., Bixler, R. D., Carraway, E. R., Moysey, S. M., Murdoch, L. R., Schlautman, M. R., & Warner, R. D. (2010). Institutionalizing undergraduate research for geology majors through creative inquiry experiences. *Journal of College Science Teaching, 39*(6), 30-37.

Wei, C. A., & Woodin, T. (2011). Undergraduate research experiences in biology: Altenatives to the apprenticeship model. *CBE-Life Science Education, 10*(2), 123-131.

Weston, T. J., & Laursen, S. L. (2015). The Undergraduate Research Student Self-Assessment (URSSA): Validation for use in program evaluation. CBE-Life Sciences Education, 14(3), ar33.

Wheaton, B., Muthén, B., Alwin, D. F., & Summers, G. F. (1977). Assessing reliability and stability in panel models. In D. R. Heise (Ed.), *Sociology Methodology* (pp. 84-136). San Francisco, CA: Jossey-Bass.

Willis, D. A., Krueger, P. S., & Kendrick, A. (2013). The influence of a research experiences for undergraduates program on student perceptions and desire to attend graduate school. *Journal of STEM Education: Innovations and Research, 14*(2), 21-28.

Wilson, A., Howitt, S., Wilson, K., & Roberts, P. (2012). Academics' perceptions of the
purpose of undergraduate research experiecnes in a research-intensive degree.
Studies in Higher Education, 37(5), 513-526.

Wong, S. L., & Hodson, D. (2009). From the horse's mouth: What scientists say about
scientific investigation and scientific knowledge. *Science Education, 93*, 109-130.

Woodzicka, J. A., Ford, T. E., Caudill, A., & Ohanmamooreni, A. (2015). A successful
model of collabortive undergraduate research: A multi aaculty, multi-project,
multi-institution team appraoch. *Teaching of Psychology, 42*(1), 60-63.

Yaffe, K., Bender, C., & Sechrest, L. (2014). How does undergraduate research
experience impact career trajectories and level of career satisfaction: A
comparative survey. *Journal of College Science Teaching, 44*(1), 25-33.

Zainudin, A. (2012). *Research methodology and data analysis* (5th ed.). Shah Alam,
Malaysia: University Technology MARA Publication Centre (UiTM Press).

Zhan, W. (2014). Research experience for undergraduate students and its impact on
STEM education. *Journal of STEM Education: Innovations and Research, 15*(1),
34-38.

Zheng, R. (2010). Effects of situated learning on students' knowledge acquisition: An
individual differences perspective. *Journal of Educational Computing Research,
43*(4), 467-487.

Zimbardi, K., Bugarcic, A., Colthorpe, K., Good, J. P., & Lluka, L. (2013). A set of
vertically integrated inquiry-based practical curricula that develop scientific
thinking skills for large cohorts of undergraduate students. *Adv Physiol Educ,
37*(4), 303-315.

APPENDICES

Note. The formatting of appendices was changed to fit in this dissertation.

Appendix A. A Sample of Data Extraction of Systematic Literature Review

Paper Tille	Authors	Yeahr	Type	Research Designer	Authenticity of Research Context	Task	Impact	sample size	Study design	Type	Valid Instru	Descriptive/Inferential	Impact Yes	No	Weight of Evidence	Note	
A course-based undergraduate	Shanle, Tsun, & Strahl	2016	CB	F	Yes	Yes	Yes	16	Survey, Instructor opinion-based	NFE	N		confidence in use tools and techniques. develop a	Identification as a scientist	4	No infor about survey	
A Perspective of Gender Differences	Harsh, Maltese, &	2012	LS	NA	NA	NA	NA	116 individual	Survey & Interview	M	No Infor	D	Higher effect of UREs on female than male students in		2		
A set of vertically integrated inquiry-	Zimbard, Bugarcic,	2012	CB	Course 1 and 2,	N	N	N	149/775 course	Course evaluation and survey	M	ratings on	I-t-test of pre-post	scientific writing, hypthesis design, experimental design,		2		
A Successful Model of Collaborative	Woodzicha, Ford, Caudill,	2015	CB	S	Yes	Yes	N	22(12 multiple,	Survey (Comparison)	QA	Y	t-test of pre-post	Experimental group-fuller appreciation of research as a	Research skills	3	Comparison study	
Academic Factors That Affect	Taraban & Logue	2012	LS	NA	NA	NA	NA	353 (59.1%)	Validated Survey, GPA, Correlations,	QA	Y	D&I correlation	Revealed significant predictors of URQ factors:		1	correlation	
Academics' perceptions of the	Wilsona, Howittc,	2012	CB	NA	NA	NA	NA	87 academic	Survey and open-ended questions.	M	N	D	Retention or selection; Research exposure;		3		
An Example of Practical	Ellis-Monaghan &	2013	SS	F	Yes	Yes	yes		Students reflections	QL			experienced scientific inquiry process.		3		
Bridging the Undergraduate	Russell, D'Costa,	2015	CB	F	Yes	Yes	Yes		pre- and post exposure tests,	QA	N	t-test of two groups pre-post	Content comprehension, confidence.		2	Control study	only post
Course-Integrated Undergraduate	Nadelson, Walters, &	2010	CB	Mixed	NA	NA	NA	17-A; 25-B; 13-C	Survey and open-ended questions.	M	N	ANOVA Post test-	Interest, engagement, learn content, Level 3 inquiry		3	comparison	only post
Demonstrated Successful	Culp & Urtel	2013	SS	F	Yes	Yes	Yes	NA	Instructor opinion-based commentary	NFE			conferences, publications, confidence.		4		
Developing Research	Davidson & Palermo	2015	CB	S	NA	NA	N	17/55	Course grade, Survey	QA	NA	t-test of pre-post	Research skills,	No satisfaction	3	Hands-on learning	pre-post
Embedding Research in	Jansen, Jadack,	2015	ND	F	Yes	Yes	Yes	NA	Observation, within project results.	NFE			experienced scientific inquiry process. Content		4		
Engaging Women In Computer Science	Kim, Fann, & Misa-	2011	ND	NA	NA	NA	NA	117 NSF funded	Surveys and interviews.	M	N	D	This study found that women benefit from participating in		2	Gender effect	
How Does Undergraduate	Yaffe, Bender, &	2014	ND	NA	NA	NA	NA	423 survey, 23	Survey and interview	M	N	D	fosters identification with the institution; research		3	Comparison, but no t-test	

Appendix B. Student Science Learning Gains Survey

Dear Student,

We are assessing how your participation in this course affects your science learning competency and gains. Participation should take about five minutes and will provide valuable information to help us improve your classroom experience and our program.

This is the first of a two part survey; the second part of the survey will be given at the end of the course. If you prefer not to use your real name, please use the same name in both surveys.

All information provided is confidential and will be used in aggregate form, your identity will remain anonymous. Your participation in this survey is voluntary and you may "opt out" at any time. There is no penalty to you for opting out - your grade for the course will not be affected. There is no personal risk or benefit associated with your participation in this survey.

If you have questions or concerns, please contact Katey Houmiel at houmik@spu.edu.

Thank you very much for your help!
The GENI Assessment Team

IRB #121306008. Expiration 1.4.2017

* Required

<div align="center">Top of Form</div>

Your School Name:

Your Course Name:

Your Name: *

(if you choose to use a pseudonym, please use the same name on both the Pre and Post Surveys)

Your gender:

- ○ Male
- ○ Female

Your major

Your year:

- ○ Freshman
- ○ Sophomore
- ○ Junior
- ○ Senior
- ○ Other: [   ]

Are you Hispanic or Latino?

- ○ Yes
- ○ No
- ○ I prefer not to respond
- ○ Other: [   ]

Which of the following best describes your Hispanic origin or descent?

- ○ Mexican or Chicano
- ○ Puerto Rican
- ○ Cuban
- ○ I prefer not to respond
- ○ Other: [   ]

What is your racial Background? Check one or more boxes

- ○ Native Hawaiian or other Pacific Islander
- ○ Asian
- ○ Black or African - American
- ○ White
- ○ I prefer not to respond
- ○ American Indian or Alaska Native (Please specify tribal affiliation in the "other" option.
- ○ Other: [   ]

How many college level science courses have you taken? *

[   ]

How many college level science courses have you taken in which you have participated in an authentic research project?

(i.e - A course into which an original research project has been integrated.)

Upon graduation I plan to pursue:

(STEM = Biology, Chemistry, Engineering, Mathematics, Physics)

- ○ a Masters degree program in the STEM sciences.
- ○ a Doctoral degree program in the STEM sciences.
- ○ a job in the STEM sciences.
- ○ a job unrelated to the STEM sciences.
- ○ postgraduate studies in the Professional Health Sciences Field (medicine, dentistry, PT, etc)

Why did you decide to take this course?

- ○ To fulfill a requirement for my major.
- ○ Because it is important for graduate or professional school.
- ○ Because it is important for my desired employment.
- ○ Because I am interested in the subject matter.
- ○ To learn laboratory skills & techniques.
- ○ To learn about the research process.
- ○ To get "hands-on" research experience.
- ○ Because the course has a good reputation.
- ○ Because the instructor has a good reputation.
- ○ Other:

1.1 My participation in authentic research will be interesting, enhanced my learning and will allow me to contribute to the scientific knowledge base.

1=not at all, 2=just a little, 3=somewhat, 4= a lot, 5=a great deal

      1   2   3   4   5

Not at all ○   ○   ○   ○   ○   A great deal

1.2 I am enthusiastic about participating in more authentic research projects integrated in science courses (when applicable).

1=not at all, 2=just a little, 3=somewhat, 4= a lot, 5=a great deal

      1   2   3   4   5

Not at all ○   ○   ○   ○   ○   A great deal

1.3 I expect that my learning experience in this course will facilitate my continuing education in the sciences, my career, and/or my life.

1=not at all, 2=just a little, 3=somewhat, 4= a lot, 5=a great deal

        1   2   3   4   5

Not at all ○   ○   ○   ○   ○   A great deal

1.4 I am confident that I can overcome obstacles encountered in the laboratory and acquire accurate and reliable results from my work.

1=not at all, 2=just a little, 3=somewhat, 4= a lot, 5=a great deal

        1   2   3   4   5

Not at all ○   ○   ○   ○   ○   A great deal

1.5 I have a well-defined problem-solving strategy for identifying critical resources and methods that I can use to more fully understand the classroom and laboratory components of this course.

1=not at all, 2=just a little, 3=somewhat, 4= a lot, 5=a great deal

        1   2   3   4   5

Not at all ○   ○   ○   ○   ○   A great deal

1.6 At the beginning of this course, I am confident that if I work hard and persist, I can attain quality results from my research.

1=not at all, 2=just a little, 3=somewhat, 4= a lot, 5=a great deal

        1   2   3   4   5

Not at all ○   ○   ○   ○   ○   A great deal

1.7 I see myself as a part of the intellectual effort in our group research project rather than as an assistant.

1=not at all, 2=just a little, 3=somewhat, 4= a lot, 5=a great deal

1   2   3   4   5

_____

Not at all ○   ○   ○   ○   ○   A great deal

_____

1.8 I am comfortable discussing complex scientific ideas and questions.

1=not at all, 2=just a little, 3=somewhat, 4= a lot, 5=a great deal

1   2   3   4   5

_____

Not at all ○   ○   ○   ○   ○   A great deal

_____

2.1 I understand the main concepts taught in this course:

1=not at all, 2=just a little, 3=somewhat, 4= a lot, 5=a great deal

1   2   3   4   5

_____

Not at all ○   ○   ○   ○   ○   A great deal

_____

2.2 I can provide examples of how the main concepts taught in this course relate to each other.

1=not at all, 2=just a little, 3=somewhat, 4= a lot, 5=a great deal

1   2   3   4   5

_____

Not at all ○   ○   ○   ○   ○   A great deal

_____

2.3 I can provide examples of how ideas taught in this course relate to those taught in other courses I have taken.

1=not at all, 2=just a little, 3=somewhat, 4= a lot, 5=a great deal

1   2   3   4   5

_____

Not at all ○   ○   ○   ○   ○   A great deal

_____

2.4 I can explain how my experience in this course might impact my thinking about issues I encounter in my everyday life (e.g. society and personal health).

1=not at all, 2=just a little, 3=somewhat, 4= a lot, 5=a great deal

1   2   3   4   5

Not at all ○ ○ ○ ○ ○ A great deal

3.1 I can follow a detailed scientific protocol.

1=not at all, 2=just a little, 3=somewhat, 4= a lot, 5=a great deal

1   2   3   4   5

Not at all ○ ○ ○ ○ ○ A great deal

3.2 I can pose a hypothesis or troubleshoot a protocol based on observations I make in the laboratory.

1=not at all, 2=just a little, 3=somewhat, 4= a lot, 5=a great deal

1   2   3   4   5

Not at all ○ ○ ○ ○ ○ A great deal

3.3 I know how to use instrumentation and laboratory techniques I expect to use in this course.

1=not at all, 2=just a little, 3=somewhat, 4= a lot, 5=a great deal

1   2   3   4   5

Not at all ○ ○ ○ ○ ○ A great deal

3.4 I can identify possible variables and problems that may influence the experiment or the operation of the equipment.

1=not at all, 2=just a little, 3=somewhat, 4= a lot, 5=a great deal

1   2   3   4   5

Not at all ○ ○ ○ ○ ○ A great deal

3.5 I can refine and modify experimental steps based on observations and outcomes from the preceding experiment.

1=not at all, 2=just a little, 3=somewhat, 4= a lot, 5=a great deal

1   2   3   4   5

Not at all ○ ○ ○ ○ ○ A great deal

3.6 I can carefully observe and record results of an experiment.

1=not at all, 2=just a little, 3=somewhat, 4= a lot, 5=a great deal

1   2   3   4   5

Not at all ○ ○ ○ ○ ○ A great deal

3.7 I can recognize quality results among the combined data I collect in the lab.

1=not at all, 2=just a little, 3=somewhat, 4= a lot, 5=a great deal

1   2   3   4   5

Not at all ○ ○ ○ ○ ○ A great deal

3.8 I can explain and synthesize experimental results into a coherent conclusion.

1=not at all, 2=just a little, 3=somewhat, 4= a lot, 5=a great deal

1   2   3   4   5

Not at all ○ ○ ○ ○ ○ A great deal

3.9 I can find and critically read scientific papers, manuals related to laboratory procedures, and relevant and reliable internet resources.

1=not at all, 2=just a little, 3=somewhat, 4= a lot, 5=a great deal

1   2   3   4   5

Not at all ○ ○ ○ ○ ○ A great deal

3.10 I can recognize sound scientific argument (or sound application of scientific technique) and appropriate use of evidence.

1=not at all, 2=just a little, 3=somewhat, 4= a lot, 5=a great deal

1   2   3   4   5

Not at all ○ ○ ○ ○ ○ A great deal

3.11 I can write research documents or give research presentations in a discipline-appropriate style and format.

1=not at all, 2=just a little, 3=somewhat, 4= a lot, 5=a great deal

      1   2   3   4   5

Not at all ○   ○   ○   ○   ○   A great deal

3.12 I can present and discuss my data and results using graphs or mathematical relationships where appropriate.

1=not at all, 2=just a little, 3=somewhat, 4= a lot, 5=a great deal

      1   2   3   4   5

Not at all ○   ○   ○   ○   ○   A great deal

3.13 I can work effectively with others, including coordinating activities, sharing my opinions, and discussing results with my peers.

1=not at all, 2=just a little, 3=somewhat, 4= a lot, 5=a great deal

      1   2   3   4   5

Not at all ○   ○   ○   ○   ○   A great deal

3.14 I can compare data collected from multiple experiments, instruments, or types of analyses.

1=not at all, 2=just a little, 3=somewhat, 4= a lot, 5=a great deal

      1   2   3   4   5

Not at all ○   ○   ○   ○   ○   A great deal

4.1 I use systematic reasoning in my approach to solving problems and can describe this approach to others.

1=not at all, 2=just a little, 3=somewhat, 4= a lot, 5=a great deal

      1   2   3   4   5

Not at all ○   ○   ○   ○   ○   A great deal

4.2 I can identify specific instances where I have applied what I learned in my science classes to situations I've encountered outside the classroom.

1=not at all, 2=just a little, 3=somewhat, 4= a lot, 5=a great deal

      1   2   3   4   5

Not at all ○   ○   ○   ○   ○ A great deal

4.3 I can provide specific examples from my life outside of school where I have used a critical approach to analyze data or develop arguments.

1=not at all, 2=just a little, 3=somewhat, 4= a lot, 5=a great deal

      1   2   3   4   5

Not at all ○   ○   ○   ○   ○ A great deal

Bottom of Form

Appendix C. Test of Student Scientific Literacy Skills

Dear Student,

We are assessing how your participation in this course affects your scientific literacy skills. Your participation will provide valuable information that will help us improve your classroom experience and our program.

There are 28 multiple-choice questions and the test should take no longer than 35 minutes.

This is part one of a two part survey; the second part of the survey will be given at the end of the course. If you prefer not to include your real name, please use the same name in both surveys. All information provided is confidential and will be used in aggregate form.

Thank you very much for your help!

* Required

Your School Name:

Your Course Name:

Your Name:

(if you choose to use a pseudonym, please use the same name on both the Pre and Post Tests)

Your Gender:

o ○ Male
o ○ Female

Your Major:

Your Year:

o ○ Freshman
o ○ Sophomore
o ○ Junior
o ○ Senior

- ○ Other: [ ]

Are you Hispanic or Latino?

- ○ Yes
- ○ No
- ○ I prefer not to respond
- ○ Other: [ ]

Which of the following best describes your Hispanic origin or descent?

- ○ Mexican or Chicano
- ○ Puerto Rican
- ○ Cuban
- ○ I prefer not to respond
- ○ Other: [ ]

What is your racial Background? Check one or more boxes

- ○ Native Hawaiian or other Pacific Islander
- ○ Asian
- ○ Black or African - American
- ○ White
- ○ I prefer not to respond
- ○ American Indian or Alaska Native (Please specify tribal affiliation in the "other" option.
- ○ Other: [ ]

How many college level science courses have you taken (Biology, Chemistry, Physics)? *

[ ]

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.

o ◯ 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?

o ◯ A

o ◯ B

o ◯ C

o ◯ D

Time (days)	Insect Population (number)
2	7
4	16
8	60
10	123
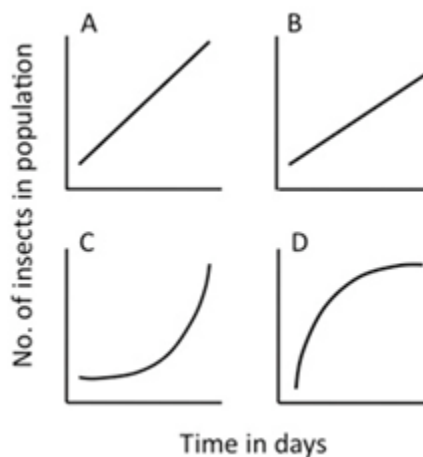
3. A study about life expectancy was conducted using a random sample of 1,000 participants from the United States. In this sample, the average life expectancy was 80.1 years for females and 74.9 years for males. What is one way that you can increase your certainty that women truly live longer than men in the United States' general population?

o ◯ a. Subtract the average male life expectancy from the average female expectancy. If the value is positive, females live longer.

o ◯ b. Conduct a statistical analysis to determine if females live significantly longer than males.

o ◯ c. Graph the mean (average) life expectancy values of females and males and visually analyze the data.

o ◯ d. There is no way to increase your certainty that there is a difference between sexes.

4. Which of the following research studies is least likely to contain a confounding factor (variable that provides an alternative explanation for results) in its design?

- ○ a. Researchers randomly assign participants to experimental and control groups. Females make up 35% of the experimental group and 75% of the control group.
- ○ b. To explore trends in the spiritual/religious beliefs of students attending U.S. universities, researchers survey a random selection of 500 freshmen at a small private university in the South.
- ○ c. To evaluate the effect of a new diet program, researchers compare weight loss between participants randomly assigned to treatment (diet) and control (no diet) groups, while controlling for average daily exercise and pre-diet weight.
- ○ d. Researchers tested the effectiveness of a new tree fertilizer on 10,000 saplings. Saplings in the control group (no fertilizer) were tested in the fall, whereas the treatment group (fertilizer) were tested the following spring.

5. Which of the following actions is a valid scientific course of action?

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

Background for question 6: The graph appeared in a scientific article 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).
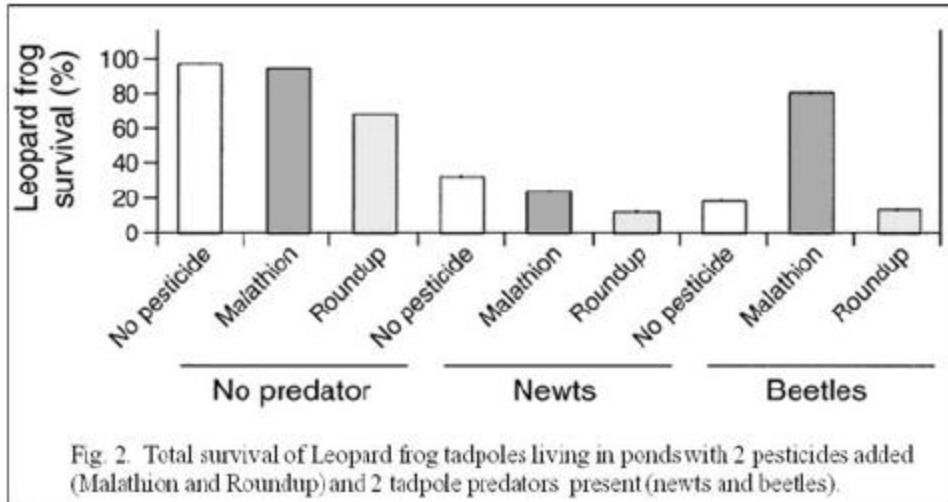
6. When beetles were introduced as predators to the Leopard frog tadpoles, and the pesticide Malathion was added, the results were unusual. Which of the following is a plausible hypothesis to explain these results?

o   a. The Malathion killed the tadpoles, causing the beetles to be hungrier and eat more tadpoles.

o   b. The Malathion killed the tadpoles, so the beetles had more food and their population increased.

o   c. The Malathion killed the beetles, causing fewer tadpoles to be eaten.

o   d. The Malathion killed the beetles, causing the tadpole population to prey on each other.

7. Which of the following is the best interpretation of the graph below?

o   a. Type "A" mice with Lymphoma were more common than type "A" mice with no tumors.

o   b. Type "B" mice were more likely to have tumors than type "A" mice.

o   c. Lymphoma was equally common among type "A" and type "B" mice.

o   d. Carcinoma was less common than Lymphoma only in type "B" mice.

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.
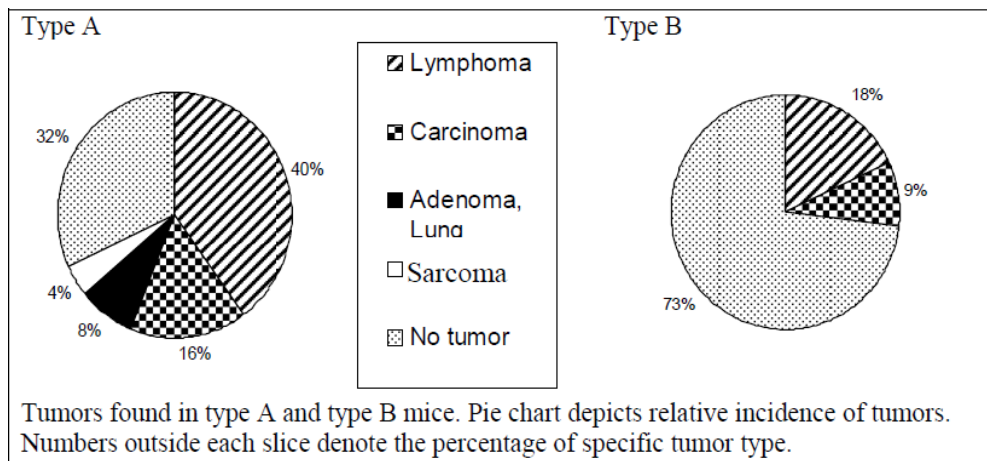
8. Creators of the Shake Weight, a moving dumbbell, claim that their product can produce "incredible strength!" Which of the additional information below would provide the strongest evidence supporting the effectiveness of the Shake Weight for increasing muscle strength?

- ○ 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.

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 most important in your confidence that the resource is accurate or not.

- ○ a. The resource may not be accurate, because appropriate references are not provided.
- ○ b. The resource may not be accurate, because the purpose of the site is to advertise a product.
- ○ c. The resource is likely accurate, because appropriate references are provided.
- ○ d. The resource is likely accurate, because the website's author is reputable.

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

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

11. The findings of this study suggest that consuming diet soda might lead to increased risk for heart attacks and strokes. From the statements below, identify additional evidence that supports this claim:

- a. Findings from an epidemiological study suggest that NYC residents are 6.8 times more likely to die of vascular-related diseases compared to people living in other U.S. cities.
- b. Results from an experimental study demonstrated that individuals randomly assigned to consume one diet soda each day were twice as likely to have a stroke compared to those assigned to drink one regular soda each day.
- c. Animal studies suggest a link between vascular disease and consumption of caramel-containing products (ingredient that gives sodas their dark color).
- d. Survey results indicate that people who drink one or more diet soda each day smoke more frequently than people who drink no diet soda, leading to increases in vascular events.

12. The excerpt above comes from what type of source of information?

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

13. The lead researcher was quoted as saying, "I think diet soda drinkers need to stay tuned, but I don't think that anyone should change their behaviors quite yet." Why didn't she warn people to stop drinking diet soda right away?

- a. The results should be replicated with a sample more representative of the U.S. population.
- b. There may be significant confounds present (alternative explanations for the relationship between diet sodas and vascular disease).
- c. Subjects were not randomly assigned to treatment and control groups.
- d. All of the above

14. Which of the following attributes is not a strength of the study's research design?"

- a. Collecting data from a large sample size.
- b. Randomly sampling NYC residents.
- c. Randomly assigning participants to control and experimental groups.
- d. All of the above

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

- Graph A
- Graph B
- Graph C
- Graph D
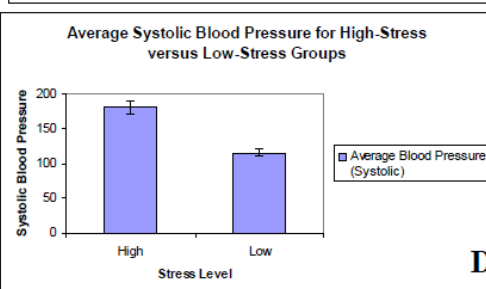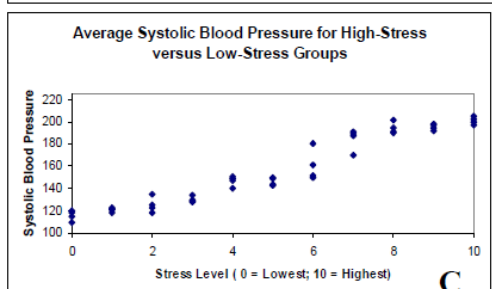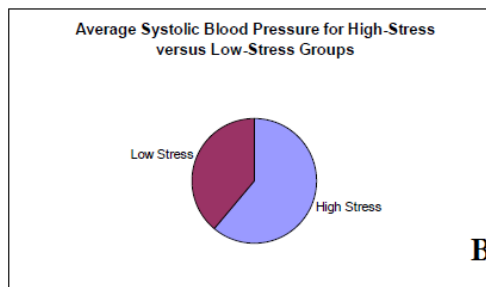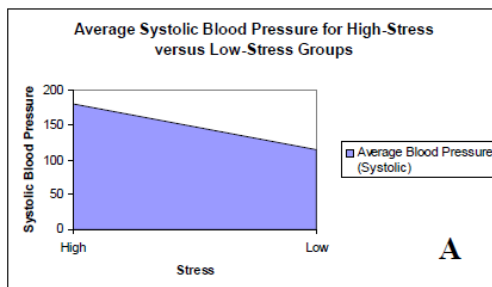
Background for question 16: Energy efficiency of houses depends on the construction materials used and how they are suited to different climates. Data was collected about the types of building materials used in house construction (results shown below). Stone houses are more energy efficient, but to determine if that efficiency depends on roof style, data was also collected on the percentage of stone houses that had either shingles or a metal roof.
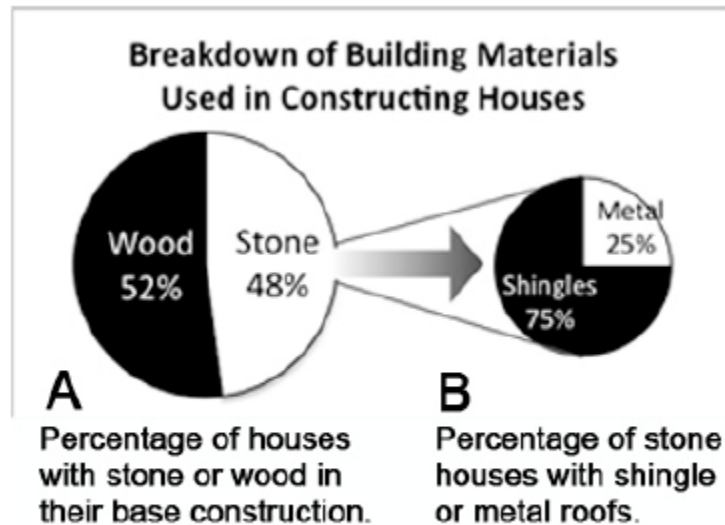
**16. What proportion of houses were constructed of a stone base with a shingled roof?**

○ ○ a. 25%

○ ○ b. 36%

○ ○ c. 48%

○ ○ d. Cannot be calculated without knowing the original number of survey participants.

17. The most important factor influencing you to categorize a research article as trustworthy science is:

○ ○ a. the presence of data or graphs

○ ○ b. the article was evaluated by unbiased third-party experts

○ ○ c. the reputation of the researchers

○ ○ d. the publisher of the article

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

○ ○ 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.

Meat consumption in developing countries
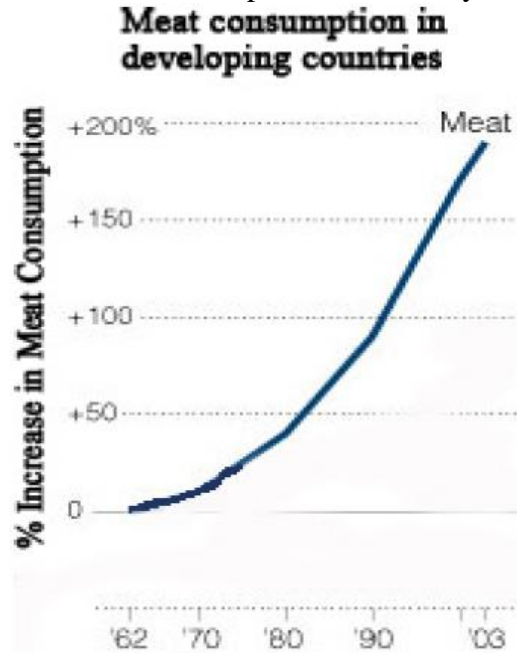
19. Two studies estimate the mean caffeine content of an energy drink. Each study uses the same test on a random sample of the energy drink. Study 1 uses 25 bottles, and study 2 uses 100 bottles. Which statement is true?

○ ○ a. The estimate of the actual mean caffeine content from each study will be equally uncertain.

○ ○ b. The uncertainty in the estimate of the actual mean caffeine content will be smaller in study 1 than in study 2.

○ ○ c. The uncertainty in the estimate of the actual mean caffeine content will be larger in study 1 than in study 2.

○ ○ d. None of the above

20. A hurricane wiped out 40% of the wild rats in a coastal city. Then, a disease spread through stagnant water killing 20% of the rats that survived the hurricane. What percentage of the original population of rats is left after these 2 events?

○ ○ a. 40%

○ ○ b. 48%

○ ○ c. 60%

o   d. Cannot be calculated without knowing the original number of rats.

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 (please see 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.

Violent crime rates
Adjusted victimization rate per 1,000 persons age 12 and over
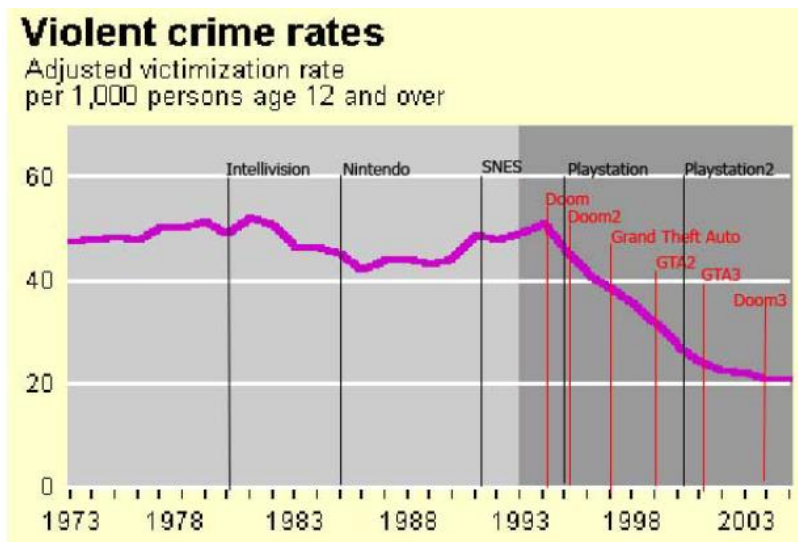
21. Considering the information presented in this graph, what is the most critical flaw in the blogger's argument?

o   a. Violent crime rates appear to increase slightly after the introduction of the Intellivision and SNES game systems.

o   b. The graph does not show violent crime rates for children under the age of 12, so results are biased.

o   c. The decreasing trend in violent crime rates may be caused by something other than violent video games

o   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?

o   a. the drug manufacturer's pamphlet/website

o   b. a special feature about the drug on the nightly news

- o c. a research study conducted by outside researchers
- o 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?

- o a. 5
- o b. 35
- o c. 50
- o d. 500

24. Why do researchers use statistics to draw conclusions about their data?

- o a. Researchers usually collect data (information) about everyone/everything in the population.
- o b. The public is easily persuaded by numbers and statistics.
- o c. The true answers to researchers' questions can only be revealed through statistical analyses.
- o 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?

- o a. a count of the number of children who were immunized and have autism
- o b. yearly screening data on autism symptoms for immunized and non-immunized children from birth to age 12
- o c. mean (average) rate of autism for children born in the United States
- o d. mean (average) blood mercury concentration in children with autism

26. Pick the best answer that would help you decide about the credibility of the Eurasian Journal of Bone and Joint Medicine:

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.

- o a. It is not a credible source of scientific research because there were advertisements within the journal.
- o 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.
- o 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.
- o 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?

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

Background for question 28: Researchers interested in the relation between River Shrimp (Macrobrachium) abundance and pool site elevation, presented the data in the graph (Please see the graph below. Interestingly, the researchers also noted that water pools tended to be shallower at higher elevations.
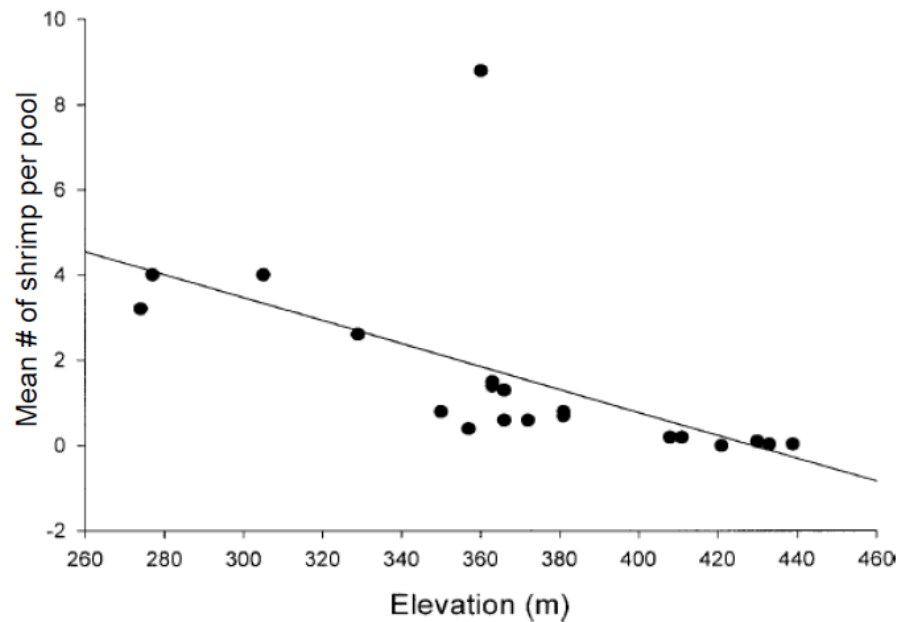
FIG. 3. Relationship between total abundance of *Macrobrachium* (1988–2002) and elevation in Quebrada Prieta.

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

o ○ a. There are more water pools at elevations above 340 meters because it rains more frequently in higher elevations.

o ○ b. River shrimp are more abundant in lower elevations because pools at these sites tend to be deeper.

o ○ c. This graph cannot be interpreted due to an outlying data point.

o ○ d. As elevation increases, shrimp abundance increases because they have fewer predators at higher elevations.