THE RELATIONSHIP OF FORMAL REASONING, MOTIVATION,

AND CONCEPTUAL CHANGE:

A QUANTITATIVE STUDY OF INTRODUCTORY BIOLOGY STUDENTS

ACROSS THE UNITED STATES

by

Romola Alaica Bernard

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

Doctor of Philosophy

in

Education

MONTANA STATE UNIVERSITY
Bozeman, Montana

November 2018

ABSTRACT

There is a noticeable disconnect between conceptual change research carried out in different domains of knowledge. This is starkly apparent in the divide between theoretical models of conceptual change stemming from cognitive and educational psychology, and empirical studies on conceptual change rooted in science education. This study operationalized models of conceptual change that accounted for the rational aspect of conceptual change that dominates in the natural sciences, and the extrarational aspects of conceptual change that are focal in the social sciences. Mixed effects models of conceptual change were investigated. In addition to prior knowledge, formal reasoning ability was incorporated as a critical rational aspect of conceptual change. Academic motivation, plus the teaching and learning environment students experience were included as essential extrarational aspects of conceptual change. The final operational model of conceptual change has post-instruction score as the response variable, and pre-instruction score, formal reasoning ability, intrinsic motivation, representation of racial group in science, teacher experience, and teaching practice as the most important predictors of conceptual change. Prior knowledge and formal reasoning ability are by far the strongest predictors of improving post-instruction conceptual understanding of evolution by natural selection for introductory biology students. There are two noteworthy findings. One, a crucial student characteristic, formal reasoning ability, has been ignored in conceptual change research. When formal reasoning ability is included as a predictor, self-efficacy is not at all important in predicting conceptual change. Two, another student characteristic, race, plays an important role in predicting conceptual change.

DEDICATION

This has been a journey of faith, endurance, and thanksgiving.

I thank my Heavenly Father, and my Lord and Savior Jesus Christ for the blessing of love, grace, and faithfulness. For without You, I toil and labor in vein; neglecting to fulfill Your divine purpose for me. It is because of the pouring out of the Holy Spirit that I have endured and completed this task. Now, I lift my hands to You, Almighty God, in thanksgiving and praise.

To my dear family, I offer my deepest gratitude for your unfailing love and unwavering support. I especially thank my son, Lincoln; my mother, Joyce; and my grandmother, Maudlin for all you have done and continue to do to make my life's journey worthwhile.

ACKNOWLEDGEMENTS

I would like to take this opportunity to thank my chair, Dr. Carrie B. Myers, for her support, guidance, and intellectual kinship. Whether she gave me written comments, or we had an in-person meeting, or we had a conversation over the phone, I always felt I got exactly what I needed to move forward in a purposeful and focused way. Dr. Myers' in-depth, thought-provoking, and timely feedback stirred me to produce the best quality work that I could. I have learned so much from her about what it means to be a stimulating and caring mentor. I am inspired by the high standard she sets as a scholar. I hope to one day follow in her footsteps and serve my students with such integrity and compassion. I am profoundly grateful to her for stepping in when I needed someone to help me make a delicate transition in my PhD program.

I would also like to thank Dr. Steven T. Kalinowski for serving as my co-chair. I am deeply thankful for the opportunity I had to work with him on the National Science Foundation (NSF) grant and for his directorship as the principal investigator. I benefitted from our many conversations about conducting and managing a huge research project. I am very appreciative of the unique perspective our work together afforded me. I am also grateful for the financial support the grant provided. Thanks also to Dr. Mary J. Leonard for being instrumental in bringing me on board to be a part of this NSF project.

I would also like to thank Dr. Bryce E. Hughes and Dr. Sarah Schmitt-Wilson for serving on my PhD committee amidst very demanding schedules. Lastly, I want to offer a special thanks to my "research buddy" and fellow PhD candidate, Catherine M. Johnson, for her collegial support in writing this dissertation.

TABLE OF CONTENTS

TABLE OF CONTENTS CONTINUED

TABLE OF CONTENTS CONTINUED

LIST OF TABLES

LIST OF TABLES CONTINUED

LIST OF FIGURES

CHAPTER ONE

INTRODUCTION

Background

Calls to reform undergraduate science education to address concerns about the United States global competitiveness stress the importance of coherent curricula organized around core disciplinary concepts (American Association for the Advancement of Science [AAAS], 2011, 2013). Biologists have identified five core concepts that are fundamental to understanding the overarching principles essential throughout the living world. These five core concepts are (1) evolution, (2) structure and function, (3) information flow, structure, and storage, (4) pathways and transformations of energy and matter, and (5) systems (AAAS, 2011). A strong preparation in the theory of evolution is crucial for students to have a deep understanding of biological systems at all organizational levels, including molecule, cell, tissue, organ, group, population, ecosystem, and so on (AAAS, 2011). The mechanism of natural selection has served as the primary explanation of evolutionary change (Anderson, Fisher, & Norman, 2002; Nehm, Beggrow, Opfer, & Ha, 2012). But unfortunately, evolution by natural selection is generally poorly understood by students (Gregory, 2009). Students have firmly held misconceptions about evolution by natural selection that are difficult to change (Anderson et al., 2002; Bishop & Anderson, 1990; Kalinowski, Leonard, & Taper, 2016; Nehm & Schonfeld, 2008). To successfully correct their misconceptions, students must alter current knowledge that conflicts with to-be-learned concepts and undergo conceptual change to form conceptions consistent with normative scientific understanding (Chi, 2008). To further understand the role of student misconceptions in learning, this study aims to examine factors that influence conceptual change in evolutionary biology. The following sections provide background information for understanding conceptual change and sets the context for the study.

The view that students have misconceptions about certain science concepts before they are taught in a formal educational setting is not new. Nearly five decades ago, Doran (1972) developed a test to assess misconceptions students have with respect to select science concepts that potentially interfere with learning. Misconceptions are scientifically incorrect ideas commonly held by students that are persistent and robust (Leonard, Kalinowski, & Andrews, 2014). Students continue to cling to misconceptions even after instruction intended to confront and correct them (Andrews, Leonard, Colgrove, & Kalinowski, 2011; Linn, 2008). This is a clear indication that changing student misconceptions is extremely difficult. A burst of research activity on student misconceptions in the 1970s led to the early conceptual change movement of the 1980s (diSessa, 2006; Duit, Treagust, & Widodo, 2008; Vosniadou, 2008a; White & Gunstone, 2008), which posits learning is fundamentally coming to comprehend and accept ideas because they are intelligible and rational (Posner, Strike, Hewson, and Gertzog, 1982). However, as research in the field burgeoned, the position that conceptual change is also influenced by extrarational factors of students’ motivational beliefs and the context in which they learn began to take hold and gain attention (Pintrich, Marx, & Boyle, 1993; Sinatra, 2005).

According to Confrey (1990), three traditions on student conceptions in science, mathematics, and programming appear in the literature: Piaget's genetic epistemology, the philosophy of science, and systematic errors. Confrey's literature review credits Posner et al. (1982) for explicitly linking the notion of conceptual change to the philosophy of science and Piagetian theory of cognitive development. Drawing on Thomas Kuhn's (1970) description of theory change in science and borrowing terminology from Jean Piaget's (1952) theory of cognitive development, Posner et al. (1982) postulate there are two phases of conceptual change. One, *assimilation*, akin to the normal, evolutionary practice of science in which knowledge is added to the domain based on known, accepted, fundamental ideas. Two, *accommodation*, analogous to the more radical, revolutionary process in scientific discovery when basic assumptions are challenged, which must then be modified.

Adopting Piaget's (1952) constructivist perspective, Posner et al. (1982) aver paramount to the process of conceptual change is the student's existing conceptions. Faced with new knowledge, a student uses existing conceptions to ascertain the significance of the new information, its relevance, and what aspects of this new information counts as solution to a problem. This is particularly significant for the distinction Piaget (1964) makes between development of knowledge that happens spontaneously through biological maturation and learning of knowledge that occurs through manipulation of situations. Piaget defines an *operation* as a set of internal actions on the part of the learner that modifies an object of knowledge and transforms it in such a way that the learner understands how the object is constructed. These operational structures form the basis of knowledge for the four sequential stages of cognitive development (sensory-motor, pre-operational, concrete operational, and formal operational thinking) a person progresses through during a lifetime (Piaget 1964, 1972). A review of the literature on formal reasoning and science teaching offers support for the importance of Piaget's work to the advancement of learning in science (Lawson, 1985). In addition, empirical investigations provide evidence for a strong association between students' formal operational thinking skills and their ability to correctly change deeply rooted misconceptions (Kalinowski, Taper, Leonard, & Willoughby, 2018; Lawson & Thompson, 1988).

There is a debate about whether conceptual change is a uniquely rational process, or a more complex process with rational and extrarational features. On one end of the continuum is Posner et al.'s (1982) model of conceptual change with a rational aspect. Farther along the continuum is Pintrich et al.'s (1993) model of conceptual change with rational and extrarational facets. Advocating for a rational model of conceptual change, Posner et al. (1982) insist learning is a rational activity that focuses attention on what learning is and not on what factors impact learning. Making the case for a more inclusive model of conceptual change, Pintrich et al. (1993) argue extrarational qualities such as motivation are essential to learning, not just cognitive or rational factors. This contention over conceptual change brings to the fore disagreement in the scientific community about whether scientific judgments are based solely on rational factors driven by pure logic and objective findings or moderated by extrarational factors dependent on personal interest, social, and historical processes (Pintrich et al., 1993). Pintrich et al. (1993) reason the process of changing widely accepted knowledge in the scientific community may not be entirely rational, as the unsettled argument over scientific judgements implies. Therefore, arguing against a totally rational model of individual conceptual change, Pintrich et al. (1993) contend the classroom community does not operate in the purported rational manner of the scientific community. Students use logic, personal experience, and their intuition to determine what information is of value. Consequently, individual conceptual change is not *just* a rational activity. Rather, individual conceptual change is multifaceted, influenced by rational *as well as* extrarational factors. To address the inadequacies of the rational model, Pintrich et al. (1993) espouse an extended model of conceptual change predicated on cognitive, motivational, and contextual factors.

The relationships among cognitive, motivational, and contextual factors in conceptual change are interactive and dynamic, and not necessarily linear or one-to-one (Pintrich et al., 1993). In their proposed model of conceptual change, Pintrich et al. (1993) identify epistemic motivation, goals, personal interest, value beliefs, self-efficacy, and control beliefs as significant motivational factors that affect conceptual change. Additionally, Pintrich et al. (1993) stipulate the classroom environment is a critical contextual factor in the process of conceptual change. Individual characteristics and personal experiences interact with current conceptions, and these interactions govern students' readiness to consider new, disparate information and change their minds. To engage in cognitive accommodation that Posner et al. (1982) describe, students must selectively attend to new information, activate prior knowledge, integrate relevant information, and find solutions to problems. This cognitive process is complicated and likely mediated by students' personal interests, self-beliefs, and the environment in which they learn (Pintrich et al., 1993). The resultant multifaceted, theoretically complex conceptual change process is reciprocating, iterative, and gradual, occurring constantly and concurrently for every learner (Sinatra & Mason, 2008; Vosniadou, 2003).

The inability of Piaget's theory of cognitive development to explain developmental features that are not universal ushered in the debate about does the mind develop in a more *general*, unified fashion or a more *specific*, fractionated manner (Case, 1987; Flavell, 1992). Many contemporary cognitive developmentalists attempted to preserve the strengths of the classical Piagetian theory yet eliminate its weaknesses. They reasoned cognitive development has general across-domain similarities related to individuals' information-processing capacity and proficiency, while at the same time specific within-domain characteristics that depends on the content knowledge being learned (Carey & Spelke, 1994; Case, 1987; Flavell, 1992; Vosniadou & Ioannides, 1998). A domain-general cognitive developmental framework suggests individuals develop through fixed sequence of broad, across tasks-and-domain structures in which all cognitive tasks are uniform and homogeneous within a given developmental stage (akin to the Piagetian general stage-like development). On the other hand, a domain-specific schema assumes developmental acquisition proceeds independently, at variable rates and in a characteristic manner for each content area due to biological constraints that govern learning (Flavell, 1992). Advocates of the Piagetian theory of cognitive development subsequently argued an individual may function at one developmental level while learning a certain type of content, but function at another level for a different content

(Case, 1987; Flavell, 1992). Based largely on the premise that learners find some knowledge easier to acquire and understand than others, this line of argument led many cognitive developmentalists to focus on specific development within a single content area or domain of knowledge (Case, 1987; Flavell, 1992; Vosniadou, 2014). Inevitably, research on conceptual change splintered along disciplinary lines with considerable activity in physics and much less in chemistry and biology within the natural sciences (Singer, Nielsen, & Schweingruber, 2012).

Adopting a domain-specific developmental approach to the study of conceptual change at the postsecondary level, I offer the following justification for choosing biology as the domain of knowledge and the theory of evolution as the core concept to investigate. The most recent data available for the percentage of United States high school students who graduate with at least one natural science Carnegie credit show that by 2009, almost all students graduated with biology credit (96%), nearly three-quarter with chemistry credit (70%), and just over a third with physics credit (36%), respectively, (National Center for Education Statistics [NCES], 2016). Given these trends in the natural sciences, high school graduates should be reasonably prepared for introductory biology courses as they enter tertiary level institutions of learning. Within the discipline of biology, the theory of evolution is regarded as *the* central unifying theory because it explains the diversity of living organisms, and it provides a mechanism for making sense of empirical observations (Dobzhansky, 1973; Farber 2003; Urry, Cain, Wasserman, Minorsky, & Reece, 2017). Students have varying degrees of exposure to the theory of evolution by way of formal instruction, informal experiences shaped by their backgrounds, and public discourse, to name a few, prior to their enrollment in introductory biology (Nehm & Reilly, 2007; Southerland & Sinatra, 2003). Often, students form deep-seated misconceptions about the theory of evolution before they take an introductory biology course (Nehm & Reilly, 2007), and these misconceptions are remarkably resistant to change (Andrews et al., 2011).

Explanation and Definition of Study Variables

This study combines a rational and extrarational perspective to the understanding of conceptual change in evolutionary biology. The rational component of the study is centered around students' formal reasoning ability, a cognitive construct (Lawson & Thompson, 1988; Piaget 1964, 1972; Pintrich et al., 1993). The extrarational component of the study focusses on students' academic motivation, an attributional construct (Pintrich et al., 1993; Sinatra, 2005; Weiner, 1972, 1985) and the teaching practice students experience, a contextual construct (Pintrich et al., 1993; Wieman & Gilbert, 2014). Therefore, this study organizes the understanding of conceptual change around three constructs, two related to the individual learner and one related to the learning environment. This study takes a domain-specific approach to the understanding of conceptual change. A discussion follows on the role of formal reasoning ability and academic motivation in the process of individual conceptual change for the topic of evolution by natural selection in biology. After which, there is a discussion on teaching practice to promote conceptual change for the theory of evolution in biology.

Formal Reasoning Ability and Conceptual Change

Formal reasoning exemplifies a rational approach to learning and correcting misconceptions. This cognitive approach to learning characterizes logical reasoning about verbally stated hypotheses that enables understanding of abstract concepts and scientific theories (Piaget, 1964, 1972). Performance on Piagetian formal operational tasks provide a quantitative measure of formal reasoning ability that is demonstrably related to learning in science (Lawson, 1985; Lawson & Renner, 1975). For general biology content knowledge, formal reasoning ability is positively associated with academic achievement for introductory biology students (Johnson & Lawson, 1998; Lawson, Banks, & Logvin, 2007). For the topic of evolution by natural selection, formal reasoning ability is negatively related to the number of misconceptions retained after instruction for 7th grade students (Lawson & Thompson, 1988). Formal operational thinkers evaluate the shortcomings of naïve conceptions relative to the merits of evidence that support the theory of evolution by natural selection. In so doing, formal operational thinkers use deductive reasoning to reject incorrect ideas and thinking in favor of scientifically correct explanations, and thereby correct previously held misconceptions and undergo conceptual change (Lawson & Thompson, 1988). To use formal operational reasoning, students, like scientists, must evaluate the validity of arguments. They must look for counterexamples to test causal claims. They must ask what the central cause is, and what alternative causes are possible. They must grapple with how each possibility can be tested, and what is the likely outcome for each possibility. They must ask how available evidence matches expectations, and what conclusions can be drawn (Lawson, 2000b; Lawson et al., 2000). Therefore, to undergo intentional conceptual change, students must (*a*) reflect on existing knowledge, (*b*) notice difference between existing knowledge and new, contrary information, (*c*) feel discomfort at the difference, (*d*) evaluate existing knowledge and new information to decide which one to believe and accept, then (*e*) use motivated action to change or reconstruct knowledge (Hynd, 2003).

Academic Motivation and Conceptual Change

Academic motivation deals with extrarational aspects of conceptual change related to student attributes that are important to learning. Students' epistemic motivation or goals for attaining knowledge influence their quality of engagement in learning. Students bring individual interests and beliefs about what they value to the learning context, which they use to evaluate the importance and meaning of a given set of information, or tasks. Social cognitive theory emphasizes the reciprocal interactions among personal attributes, social environment, and goal-directed behaviors. It postulates students have a sense of agency that directs their actions (Bandura, 1989, 1993, 2001). For example, their personal interest to attain mastery in learning and their self-determination to exercise control over events to achieve goals (Reeve, Nix, & Hamm, 2003; Ryan & Deci, 2000; Sinatra & Mason, 2008). Students' self-efficacy, or perceptions of their ability to perform, impacts their sense of agency to initiate and sustain goal directed-behaviors (Bandura, 1989; Schunk, Meece, & Pintrich, 2014). It is prudent to make a distinction between an *intrinsic*, mastery, task-involved goal and an *extrinsic*, performance, ego-involved goal, as it pertains to social ramifications for students (Pintrich et al., 1993). Students with an intrinsic goal orientation usually strive for mastery. They typically have strong self-efficacy about completing tasks and high personal interest in those tasks. They relish the challenge and difficulty of the tasks and often deepen their engrossment in such tasks for personal satisfaction. They frequently direct their attention toward learning information related to a topic of interest for personal enjoyment (Bandura, 1994; Sinatra & Mason, 2008). On the other hand, students with an extrinsic goal orientation are usually motivated to perform for an external reward or to pursue an activity as a means towards an end. For instance, students with high self-efficacy who are confident their effort spent learning will lead to high grades and a promising career generally envision success and continue to strive, even under adverse conditions. Conversely, students with self-doubt about their capabilities often visualize failure and cease to exercise the necessary determination to attain their goal (Linnenbrink & Pintrich, 2003; Ryan & Deci, 2000; Sinatra & Mason, 2008). Cues from the social context reinforce or diminish students' view of anticipated consequences of their actions. Validation and positive feedback from peers and figures of authority fortify feelings of accomplishment, while indifference and unsupportive comments invite disposition toward defeat (Schunk et al., 2014).

Students' academic motivation may promote or impede their willingness to undergo conceptual change. Motivation to do something for its own sake is intrinsic, which is related to students' self-efficacy, self-determination, and personal interest. In contrast, motivation to attain an external outcome, such as performing well for future gains or to look good in the eyes of others, is extrinsic (Glynn & Koballa, 2006; Glynn, Taasoobshirazi, & Brickman, 2007; Reeve, Nix, & Hamm, 2003). As it concerns conceptual change, students with strong self-efficacy are more likely to feel assured in their ability to change their minds and use suitable cognitive tools to do so in the face of new knowledge that challenge their current conceptions. At the other extreme, students who are too confident in their abilities resist considering new ideas, even when their current conceptions fail to explain new information (Pintrich, 1999; Pintrich et al., 1993). Students with high self-determination who believe control over personal behavior is important when faced with evidence that call into question current understanding may contemplate new persuasive arguments that lead to a different way of thinking. However, students who are convinced that what counts as credible information are based on factors outside of their control are reluctant to examine their existing knowledge and entertain alternative points of view (Pintrich, 1999; Sinatra, Brem, & Evans, 2008). Students who are interested in a topic may be disposed to change their point of view given new information. Contrastingly, students who have a vested interest in a topic may feel compelled to remain committed to their current view (Pintrich, 1999; Sinatra & Mason, 2008). Students who work hard to get good grades and to prepare for their intended careers may be persuaded to correct their misconceptions about knowledge that may impact their future. But quite the opposite is likely for students who are less inclined to develop a deep understanding of what they are learning. These students may do only what is necessary to get by and not expend the energy needed to change any inaccurate knowledge they may currently have. The strength of students' existing ideas, the coherence of these ideas, and the commitment of students to these ideas mediate students' willingness to reexamine existing conceptions (Dole & Sinatra, 1998). The degree to which students readily scrutinize their personal beliefs and consider other perspectives is key to their ability to understand, and their willingness to eventually accept a radically different way of thinking and viewing the world (Sinatra, Southerland, McConaughy, & Demastes, 2003).

Empirical studies suggest there is a positive association between students' academic motivation and change in conceptual understanding of evolution by natural selection. For introductory biology students, self-efficacy is positively related to academic achievement (Lawson et al., 2007). For high school students, hierarchical cluster analyses indicate high personal interest and high self-efficacy paired with moderate prior knowledge yield the greatest conceptual change for girls. In comparison, moderate personal interest and moderate self-efficacy paired with high prior knowledge or high personal interest and high self-efficacy paired with moderate prior knowledge result in the greatest enduring conceptual change for boys (Linnenbrink-Garcia, Pugh, Koskey, & Stewart, 2012). These results highlight the importance of investigating different components of academic motivation, while accounting for students' prior knowledge in determining levels of conceptual change. Additionally, these results suggest it is useful to consider differences among males and females in the conceptual change process. For the two groups of students discussed, the influence of academic motivation on conceptual understanding of evolution by natural selection for high school students (mean age of 15.2) and introductory biology students (mostly freshmen) is expected to be similar because of the closeness in age of both groups and the expectation that both groups are largely at the same cognitive developmental level.

Teaching Practice and Conceptual Change

Students' prior knowledge often interferes with new to-be-learned concepts; therefore, teachers need to figure out what students already know to better help them learn challenging and counter-intuitive ideas. Drawing on their experiences teaching the theory of evolution, Sinatra et al.'s (2008) surmise students' barriers in undergoing conceptual change fall into three categories: (1) basic constraints that are present from infancy and early childhood, (2) experiences that reinforce naïve ways of thinking, and (3) emotional and motivational reactions that make it difficult to entertain the possibility of change. To increase the likelihood of effecting conceptual change, Sinatra et al. (2008) urge teachers to remember that students learning about a complex process like evolution are likely to experience conflicts, so changing students' conceptions will be difficult. Consequently, Sinatra et al. (2008) encourage teachers to prioritize trying to understand sources of students' conflicts and their resistance to accepting established principles of evolution, since this will enable teachers to design better instruction that allow students to think deeply about alternative perspectives. For instance, Kalinowski, Leonard, Andrews, and Litt, (2013) recommend teachers use activities that target students' misconceptions by posing thought-provoking questions such as, "What does natural selection predict about the evolution of antibiotic resistance in E. coli?" (p. 485), then have students assess whether the mechanism of natural selection occurs based on variation, inheritance, and differential reproductive success. In so doing, students get to formulate answers, evaluate the pros, cons, adequacy and/or inadequacy of responses with peers in small groups, and discuss their best arguments with the entire class.

Students are therefore afforded opportunities to express their current understanding publicly, get feedback from others who may or may not share their interpretation, and defend and/or assess their position, so any misconceptions they have may be exposed and addressed (Mintzes, Trowbridge, Arnaudin & Wandersee, 1991). Students generally find situations like these threatening at first. For this reason, Mintzes et al. (1991) caution teachers to be sensitive to students' vulnerability. From the very outset, teachers must foster and maintain a learning environment that encourages diverse points of view, all worthy of consideration, no matter the topic.

Statement of Problem

National reports on how to improve teaching and learning in undergraduate science make two salient recommendations that are particularly pertinent to promoting conceptual change: teaching must be based on best research practices and learning must be centered on the distinct cognitive needs and personal attributes of students (AAAS, 2011, 2013; Kober, 2015; Singer et al., 2012). While there are advances in the development of instructional strategies to effect conceptual change (Clement, 2008; Jonassen, 2008; Kalinowski et al., 2013; Linn, 2008), putting them into practice often prove problematic (Duit et al., 2008; Limón, 2001). Most of the analyses performed to evaluate the efficacy of conceptual change instructional strategies look at the learner but not at the teacher, although the teacher is integral to effectively implement any new method of teaching. One of the most common conceptual change instructional strategies attempted in the classroom is to induce cognitive conflict for students by presenting anomalous data or contradictory information. The problem with this approach is teachers are not routinely given the training they need to act as good facilitators of using conflicting evidence to help students successfully change misconceptions (Limón, 2001). Another concern is the polarity in research on conceptual change coming out of the domains of cognitive psychology and science education. Research in one domain appears to have little or no bearing on research in the other domain. As a result, discussions in cognitive psychology on different theoretical perspectives that inform a multi-dimensional approach to teaching for conceptual change seem not to reach science teachers in the classroom (Duit et al., 2008).

Another difficulty in conceptual change research is there are several leading models of conceptual change in the social science literature that differ with respect to the constructs used to describe and predict conceptual change (Murphy & Alexander, 2008). For example, recent research on conceptual change has focused considerable attention to cognitive difficulties students have in learning advanced and counter-intuitive concepts in various content areas (diSessa, 2006; Vosniadou, 2008b, 2013). However, less consideration has been given to student attributes that are critical in bringing about and sustaining long-term conceptual change (Siegel & Svetina, 2008). Given the disagreement about potential constructs that may influence the process of conceptual change and the differential attention they receive, there is a clear need for a comprehensive model of conceptual change; one that uses a rational and extrarational perspective to explain essential elements such as formal reasoning and academic motivation in the conceptual change process that is addressed in this study.

Yet another issue in conceptual change research is the gap between theoretical postulates about highly complex constructs and empirical evidence that support them (Vosniadou, 2008a). At the heart of this problem is the tension between the social and natural sciences' characterization of the word 'theory'. In the social sciences, theorists speculate about constructs and processes that are not fully or directly testable, and they propose arguments or espouse 'theory' about such phenomena (Murphy & Alexander, 2008). In contrast, in the natural sciences, 'theory' refers to a widely accepted, coherent group of propositions formulated to explain a group of facts in the natural world that are repeatedly confirmed through experiment or observation (Dictionary.com, 2018; Urry et al., 2017). This contradiction, to a large extent, speaks to the discontinuity between theoretical models originating from the social sciences and empirical studies conducted in the natural sciences on conceptual change. One meaningful way of tackling this problem is to operationalize a robust model of conceptual change that includes a rational component (e.g. formal reasoning, of interest in the natural sciences) and an extrarational component (e.g. academic motivation, widely studied in the social sciences), which is empirically tested in multiple academic settings.

A persistent focus on the efficiency of conceptual change instructional strategies compared to traditional methods favors single intervention approaches that do not necessarily lead to lasting positive changes in students' conceptions (Limón, 2001). Inconsistent results on the advantages of conceptual change instructional strategies suggest a supporting learning environment that expands the repertoire of tasks students experience, as well as augments students' interests and confidence in their capabilities are necessary and complementary for a sustained, satisfactory outcome of conceptual change (Duit et al., 2008). Therefore, to confront inadequacies in the theory and practice of conceptual change, a comprehensive model of conceptual change is needed that includes constructs relating to students' cognitive ability such as formal reasoning, personal attributes for instance academic motivation, and the educational context in which students learn. An inclusive model of conceptual change is needed that incorporates rational and extrarational constructs, which addresses criticisms of the early rational model and helps pave a path toward congruity in conceptual change research among the different domains. Also, a robust model of conceptual change is needed that is tested with data from several educational contexts to pinpoint what construct(s) is/are important in the process of conceptual change.

Statement of Purpose

The purpose of this study is to operationalize a model of conceptual change to understand the role of students' formal reasoning ability as a cognitive construct and academic motivation (namely self-efficacy, self-determination, intrinsic motivation, grade motivation, and career motivation) as an attributional construct in the conceptual change process, while accounting for the educational context students encounter as they learn. Specifically, this study uses a series of hierarchical linear modeling analyses to test the relationship of formal reasoning ability, academic motivation, and change in conceptual knowledge of evolution by natural selection, controlling for teaching practice, teacher experience, and student demographic information for introductory biology students at six postsecondary institutions across the United States.

Research Questions

To achieve its purpose, this study addresses the following research questions for undergraduate introductory biology students at six institutions for the topic of evolution by natural selection:

1. To what extent do students' formal reasoning ability and academic motivation predict levels of conceptual change?
2. Which variable, or combination of variables, that is, formal reasoning ability, academic motivation, teaching practice, teacher experience, and student demographics, are most likely to predict levels of conceptual change?

In addition to the dependent variable of conceptual change and the independent variables of students' formal reasoning ability and academic motivation, a set of control variables of students' demographic information as well as the teaching practice and teacher experience students encounter are included in the hierarchical linear modeling and hierarchical regression analyses.

Significance of Study

The findings of this study will help forge a path towards operationalizing a model of conceptual change that seeks to tackle criticisms of the early rational model by incorporating extrarational constructs important in the process of conceptual change.

This study provides empirical evidence for tangible, testable constructs that are important in the process of students changing their conceptions to be more consistent with accepted scientific knowledge. These findings may lead to the development of instructional strategies that are more accessible to and more readily adopted by classroom teachers because this study (1) targets variables to which teachers can intuitively relate, and (2) accounts for multiple educational settings arguably representative of everyday classroom experiences. Effective conceptual change instruction will enable students to understand difficult scientific ideas, better prepare for careers in science, and grapple with public issues increasingly related to science from an informed point of view. The findings of this study may inform the national conversation by highlighting the importance of students' individual characteristics to their persistence in science courses and likeliness to graduate with a degree in a science discipline.

Theoretical Framework

As mentioned previously, the goal of this study is to operationalize a model of conceptual change that incorporates cognitive, attributional, and contextual constructs to explain students' conceptual change. The explanatory constructs may be divided into three sets of variables: student variables that are assessed and measured, instructor variables that are evaluated and provided, and control variables for students that are included (Figure 1.1 on page 22). The student variables comprise the cognitive construct of formal reasoning ability and the attributional construct of academic motivation. The instructor variables deal with the contextual construct of teaching practice and the length of time teaching the specified subject-matter. The control variables consider students' demographic information relevant to learning the content knowledge, not otherwise accounted for by the model. The response construct is conceptual change of evolution by natural selection, explored further in chapter 2. What follows is a discussion on the explanatory variables.

First, the student variables are described. Students' formal reasoning ability is assessed using paper-and-pencil tests to complete Piagetian formal operational tasks of controlling variables and testing hypotheses using deductive reasoning, as well as reasoning with proportion, probability, correlation, and logic (Kalinowski et al., 2018). Also, students self-report their measure of academic motivation on self-efficacy, self-determination, intrinsic motivation, grade motivation, and career motivation scales (Glynn, Brickman, Armstrong, & Taasoobshirazi, 2011). Students' intrinsic motivation for mastery of content and personal fulfillment are influenced by their self-efficacy, self-determination, and interests. In comparison, students' extrinsic motivation to perform well and prepare for their future are linked to their grade and career motivation (Glynn & Koballa, 2006; Glynn, Taasoobshirazi, & Brickman, 2009; Linnenbrink & Pintrich, 2003; Ryan & Deci, 2000).

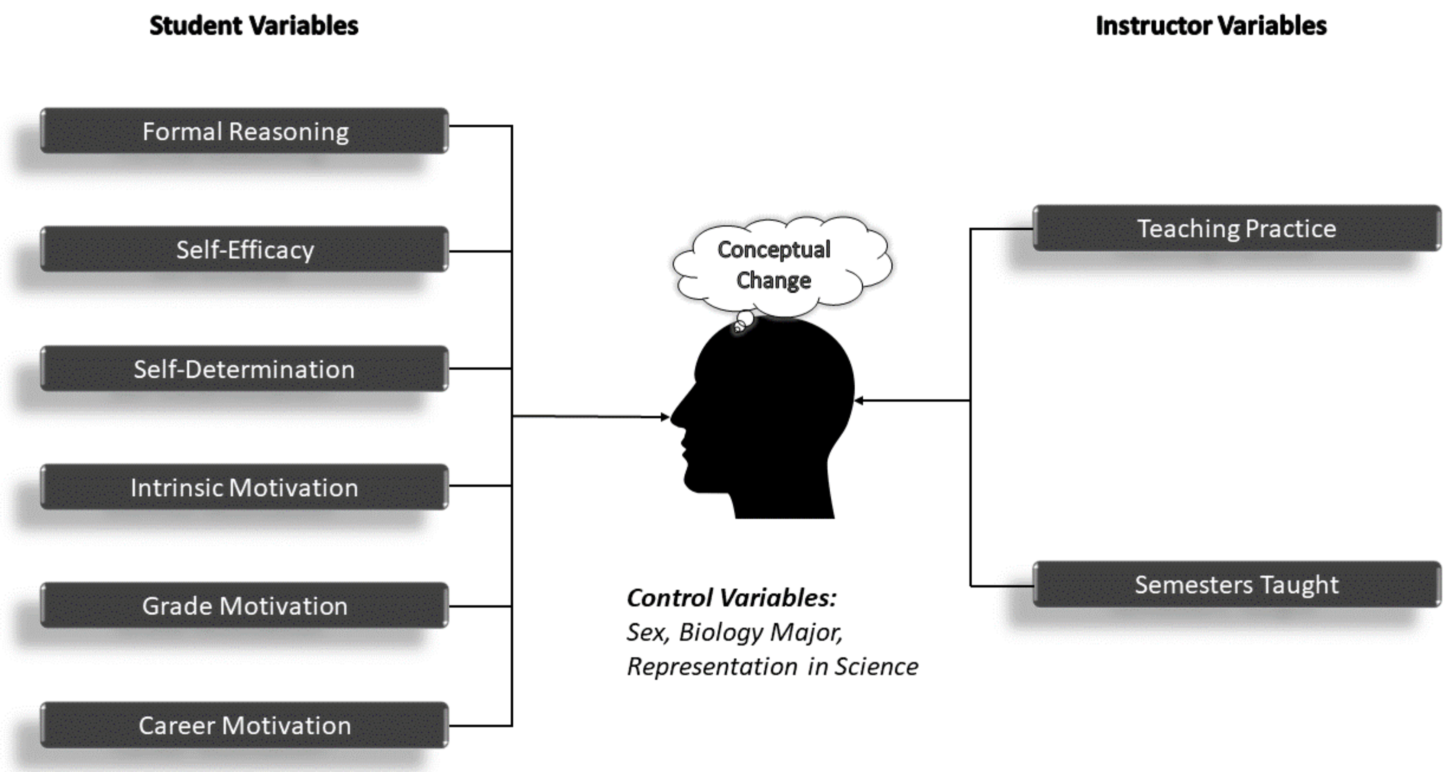


Figure 1.1. Theoretical framework illustrating the student, instructor, and control variables that influence students' conceptual change

Next, the instructor and control variables are described. Instructors of the introductory biology course complete a teaching practice inventory designed to evaluate levels of research-based postsecondary teaching in science (Wieman & Gilbert, 2014). Instructors also report the number of semesters they have taught the course as a faculty member. There is well-documented evidence teacher experience and content knowledge expertise are associated with student learning (Berry, Friedrichsen, & Loughran, 2015; Loughran, Mulhall, & Berry, 2004; Rosenblatt, 2011). For the control variables, students' demographic information, namely sex, declaration as a biology major, and the racial group to which students identify are included in the model. Age is considered an important demographic variable, but a lack of variation in students' age negated its inclusion. The term 'sex' refers to sex differences related only to physiology and anatomy. There are reported differences between males and females regarding their cognitive ability in certain content areas (Fennema & Sherman, 1977; Halpern, 2004) and their academic motivation to learn (Patrick, Ryan, & Pintrich, 1999). Declaration as a major in biology is included in the model to isolate and control for any explanatory power this variable may have on the findings from this analysis. Students take introductory science, technology, engineering, and mathematics (STEM) courses to meet degree requirements but may declare as a major in only one or none of these programs (Chen, 2013; Kokkelenberg & Sinha, 2010). A great proportion of students who declare as a major in a STEM field may not graduate with a degree in that field (Higher Education Research Institute [HERI], 2010; Kokkelenberg & Sinha, 2010; Rask, 2010).

Regardless, students may be more invested in learning difficult ideas, if at the time they are learning the subject-matter it has great significance for their future.

The representation of minority racial groups in science is impacted by the availability of role models with whom minority persons identify. The academic achievement of certain minority students in STEM introductory courses is reported to be associated with the race of the instructor (Price, 2010). This may be because students feel more empowered to approach instructors with whom they share cultural, racial, and ethnic similarities for guidance and help (Chism, 1994; Rendon, 1994). This supports the notion that students tend to have career role models whose race is the same as their own (Karunanayake & Nauta, 2004). The small proportion of minority faculty in science disciplines available to mentor minority students may translate to an underrepresentation of students, and ultimately professionals, in science (Nelson & Brammer, 2010). In STEM fields, particularly engineering at some institutions, students are disproportionately of Asian ethnicity (Chen, 2013; HERI, 2010; Kokkelenberg & Sinha, 2010). As such, students of Asian ethnicity are combined with White students as the majority group, and students of all other race and ethnicities combined as the minority group for the representation in science variable (HERI, 2010).

Research Design

This quantitative study uses a pre-test/post-test design (Bonate, 2000) to analyze data gathered Spring 2017 to operationalize a model of conceptual change. This study was part of a larger National Science Foundation funded project (NSF #1432577). Test and survey data were collected from introductory biology students from six different courses at six postsecondary institutions; that is, one course at each institution. Only one introductory biology instructor participated at each institution. If the course had multiple sections taught by more than one instructor, only students in sections taught by the participating instructor were included in the study. To account for the nested structure of the data (see Figure 3.1 in chapter 3 on page 82), a series of hierarchical linear models (Galecki & Burzykowski, 2013; Pinheiro, Bates, DebRoy, Sarkar, & R Core Team, 2018; Zuur, Ieno, Wlaker, Saveliev, & Smith, 2009) were run to find out which set of variables best explain conceptual change. The models analyzed the influence of formal reasoning ability and academic motivation on change in conceptual knowledge of evolution by natural selection, controlling for students' demographic information, teaching practice, and teacher experience in the educational setting. Hierarchical regression analyses (Akaike, 1974; Bartoń, 2018; Burnham & Anderson, 2004; Burnham, Anderson, & Huyvaert, 2011) were also carried out to determine the most parsimonious models of conceptual change for the data.

Assumptions

Several assumptions were made in this study. First, it is possible to numerically measure highly complex constructs using psychometric instruments that assess students' understanding and ask students and instructors to self-report on their beliefs and behavior. Though psychometric techniques increase the likelihood that the responses students and instructors provide are reliable and valid, these metrics are not absolute and not without limitations. At best, they are the most accurate estimation possible for the constructs being investigated (discussed more in chapter 2). Second, cognitive development occurs in sequential stages, culminating with formal operational thinking. While this study does not stipulate the precise age at which students develop formal operational thinking skills, it does assume students begin to advance to the formal operational thinking developmental stage at the secondary level of their education and this development continues at the postsecondary level. Third, by extension, this study assumes cognitive development is domain-specific because natural biological information-processing constraints make certain domains of knowledge inherently easier to learn than others. Fourth, the constructs of formal reasoning ability and academic motivation are measures of students' cognitive ability and personal attributes, respectively; whereas the construct of teaching practice is representative of the social environment in which students learn. Fifth, formal reasoning ability, academic motivation, and teaching practice are stable constructs that do not change over the course of a semester. Sixth, the quantitative analytical tools employed in this study are appropriate for the sampling procedures used, sample obtained, and data collected, and they provide realistic estimates of the relationships among the constructs under examination (elaborated further in chapter 3).

Delimitations

Certain decisions were made in planning and executing this study. The target population was restricted to public four-year colleges and universities with 10,000 or more students. Concentrating on larger institutions increased the probability of class sizes with 100 or more students. A choice was made to focus on introductory biology students and the theory of evolution with some justification to study conceptual change. Due to logistic and budgetary constraints, only one instructor and the students of that instructor participated at each institution. Although data were collected on the race with which each student identifies, students were divided into two groups (majority and minority) based on the representation of their racial group in science. In studying conceptual change, instead of focusing on *how* students restructure their conceptions in the process of conceptual change, this study examines *what* features and qualities of students are likely to bring about conceptual change. This study investigates students' formal reasoning ability as a cognitive construct and academic motivation as an attributional construct on the conceptual change of students' understanding of evolution by natural selection, controlling for teaching practice and teacher experience students have while learning.

Limitations

An attempt was made to obtain a large enough sample of instructors at different institutions for this study. Everything was done to obtain 15 or more instructors, but only six instructors agreed to participate in the study Spring 2017. The sampling procedure to procure instructors to participate was not truly random. Although instructors were randomly identified via their institution, instructors ultimately selected whether to participate in the study. These instructors may have a vested interest in the notion of conceptual change and/or the importance of evolution in teaching and learning.

Nevertheless, an argument can be made that the sample of introductory biology students obtained from the six participating institutions is representative of introductory biology students in the United States. The requirements for the quantitative analyses demanded we target a class size of 100 or more students for each instructor. Assuming this study comprise a representative sample of introductory biology students across the nation, inferences drawn from this study may be generalized to the population of introductory biology students in courses with 100 or more students in the United States. Otherwise, inferences made are limited to the sample of introductory biology students who participated in the study. There was no random assignment of students to experimental treatment groups. It is not possible to randomly assign cognitive ability and personal attributes to students or is it ethical to deliberately assign students to instructors based on teaching practice and teacher experience for the purposes of an experiment. Consequently, this study has an observational study design. We cannot make causal inferences with respect to formal reasoning ability, academic motivation, and conceptual change, but we can draw conclusions about the relationships of formal reasoning ability and academic motivation on conceptual change, controlling for teaching practice students experience. A cautionary note, this study measures conceptual change that occurs during one semester. No inferences can be made about conceptual change that occurs beyond a unit of instruction and/or a period of a semester.

Chapter Summary

There is a need for a comprehensive model of conceptual change that deals with criticisms of the early rational model and incorporates extrarational factors that are important to the process of conceptual change. A robust model of conceptual change accounts for multiple educational settings and have empirical evidence to support the constructs used to describe and predict conceptual change. Students' cognitive ability regarding the rational aspect of conceptual change and students' academic motivation relating to the extrarational component are operationalized into a model of conceptual change that controls for the context in which students learn by means of the teaching practice they experience. Hierarchical linear models are used to determine which sets of variables best explain conceptual change and hierarchical regression analyses are used to ascertain the most parsimonious models of conceptual change.

CHAPTER TWO

REVIEW OF RELATED LITERATURE

Introduction

Research on conceptual change has a relatively long history that started with scholarly articles on student conceptions in the 1970s and expanded to conceptual papers and empirical studies on the restructuring of knowledge to change misconceptions in the 1980s and beyond. From the very beginning, the domains of cognitive psychology and science education, and later social psychology played a central role in the character, scope, and direction of research in the theory and practice of conceptual change. Over time, partly because of disagreements about a cogent theory of cognitive development and divergent priorities among the different domains, a visible split emerged with regards to the emphasis and nature of research on conceptual change. Theoretical models of conceptual change stemming from the domain of cognitive psychology frequently lacked empirical data to support the ideas proposed. On the other hand, empirical studies conducted in the natural sciences generally disregarded extrarational constructs, such as motivation, hypothesized in the social sciences to be important in the process of conceptual change. This resulted in a noticeable separation in the tenor and objectives of conceptual change research among the respective domains. Commenting on conceptual change research over the years, Sinatra (2005) observed research in cognitive psychology concentrated mainly on mapping and explaining changes in the cognitive structures and knowledge representations of learners, while research in science education focused on designing instruction to foster conceptual change. Understanding the cognitive processes that operate during conceptual change and finding effective teaching strategies that help students change their conceptions to be consistent with normative scientific knowledge continues to be an important topic of research on the national stage. Therefore, it is important to bridge the gap that exists among different domains of knowledge to clear the way for even greater progress in the field.

Murphy and Alexander (2008) present an astute perspective on the divide between theoretical models and empirical studies published in the early 2000s on conceptual change. Murphy and Alexander (2008) identified 41 documents with 47 independent studies on conceptual change using the following criteria for each document: (1) have the term ‘conceptual change’ in the title, abstract, or keywords, (2) investigate a targeted construct, (3) be empirical in nature, (4) be fully retrievable with descriptive statistics and design features, (5) undergo peer review, (6) involve K-12 students’ conceptual change in total or in part, (7) incorporate some learning outcome, and (8) published in English. They found definitions of conceptual change varied markedly among the studies, with each definition determined by the theoretical model from which it was derived. Only 12 of 47 studies defined conceptual change as restructuring of or change in existing knowledge. A modest proportion, 35 of 47 studies, provided information on students’ cognitive ability. Murphy and Alexander’s (2008) review brings to the fore the obvious disconnect that exists in conceptual change research among the different domains. Moreover, this stark review raises questions about whether there is consensus on what conceptual change means, and if there is a clear direction for future inquiry in the field.

For an update on recent work in the field, I used a similar set of criteria as Murphy and Alexander (2008) to identify empirical studies on conceptual change published in the last 18 years (January 2000 to June 2018). I narrowed my focus to evolutionary biology and searched the Education Resources Information Center (ERIC), the American Psychological Association PsycINFO®, and Web of Science databases for keywords ‘conceptual change’, ‘misconceptions’, ‘biology’, and ‘evolution’. This resulted in 13 articles, 7 of which involved undergraduate students, whereas the remaining 6 focused on K-12 settings. Of the 7 articles about undergraduate students, 5 studies had a pre-test/post-test quantitative design that investigated change in knowledge from one point in time to another. The remaining 2 studies used a qualitative design that evaluated knowledge at a stated time. Murphy and Alexander (2008) argue a reasonable solution toward congruity between theoretical models and empirical evidence is to design studies that routinely incorporate constructs other than subject-matter knowledge, such as motivation and cognitive ability. Remarkably, of the 13 studies identified on evolutionary biology in the past 18 years, only 1 on eighth grade students met Murphy and Alexander’s requirement of incorporating at least one other construct besides subject-matter knowledge, in this case motivation. To further support this point, Sinatra and Mason (2008) observed ten years ago empirical studies investigating the relationships among attributional and affective constructs on change in conceptual knowledge are relatively rare.

This is testament to the fact that little has changed in the study of conceptual change. Empirical studies within the natural sciences focus almost exclusively on subject-matter knowledge, ignoring attributional and contextual constructs that theoretical models in the social sciences posit to be important to the process of conceptual change. To make the case for a comprehensive model of conceptual change that incorporates cognitive, attributional, and contextual constructs, first, I discuss leading theoretical models of conceptual change in the literature. Second, for the most part, I present a model of conceptual change that operationalizes formal reasoning ability as the cognitive construct, academic motivation as the attributional construct, and teaching practice as the contextual construct to describe and explain the conceptual change construct of evolution by natural selection. Figure 2.1 (page 34) shows a graphical depiction of the relationships among the related literature under review for this study.

Only peer reviewed conceptual, empirical, and review articles are considered in this critical review of related literature. Specifically, conceptual papers that present theoretical models of conceptual change from cognitive psychology, science education, educational psychology, and social psychology are critiqued. Also, empirical studies that test theoretical models of conceptual change and investigate the role of formal reasoning, academic motivation, or conceptual change for evolution by natural selection in learning are reviewed. In addition, empirical studies on teaching practice in science education are perused. Finally, scholarly articles that review or discuss the topics of conceptual change for evolution by natural selection, formal reasoning, academic motivation, and teaching practice are also examined.

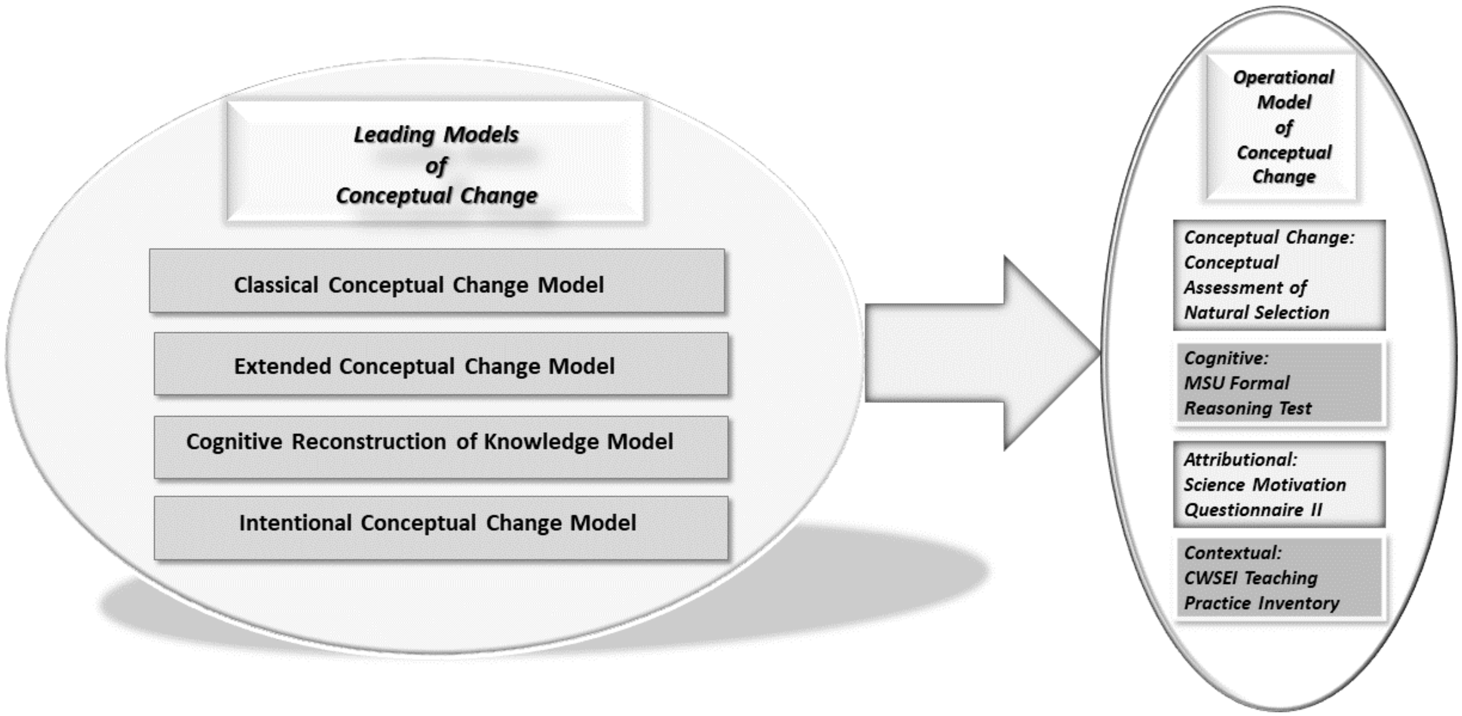


Figure 2.1. Structure of related literature for study

Leading Models of Conceptual Change

The decades of the 1980s and 1990s witnessed the most insightful and influential scholarly publications in the field of conceptual change. This period, followed by the early 2000s, marked an enormously productive era. The profound impact of seminal articles on models of conceptual change published during this fruitful time still reverberates today, some 25 to 35 years later. One of the most frequently cited work is Posner et al.'s (1982) classical conceptual change model (Google Scholar search carried out on July 15, 2018 reported 6,785 citations). The classical conceptual change model, with its rational perspective, is considered the benchmark as it is the first notable articulation of what conceptual change learning entails (Vosniadou, 2003, 2008a). Notwithstanding, the publication of Pintrich et al.'s (1993) extended conceptual change model ushered in a turning point in conceptual change research because this model integrated rational and extrarational aspects to the study of conceptual change. Its prominence in the field is evidenced by 2,756 citations (Google Scholar search on July 15, 2018), and its role in paving the path for a legacy of outstanding contributions to the study of conceptual change (Sinatra, 2005). Dole and Sinatra's (1998) cognitive reconstruction of knowledge model signaled another important juncture in conceptual change research. This model embodies the skillful and deliberate amalgamation of research from cognitive psychology, science education, and social psychology that elaborated on rational and especially extrarational facets of the conceptual change process (745 citations from Google Scholar search on July 15, 2018). The treatise on intentional conceptual change edited by Gale M. Sinatra and Paul R. Pintrich comprise a reflection on and contemplation of current understanding and future directions for conceptual change research by leading scholars in the field (Sinatra & Pintrich, 2003a, 2003b). This book, with 402 citations (Google Scholar search on July 15, 2018), documents the early developments of one of the most recent models of conceptual change. In the following subsections, each of these four leading models of conceptual change (Posner et al., 1982; Pintrich et al., 1993; Dole & Sinatra, 1998; Sinatra & Pintrich, 2003a, 2003b) is critiqued for its unique contribution to the advancement of our understanding of the theory and practice of conceptual change.

Classical Conceptual Change Model

The model of conceptual change advanced by Posner et al. (1982) is widely regarded as the classical conceptual change model that has and continues to direct research and practice in science education (Sinatra, 2005; Vosniadou, 2003). Shaped by Piaget's constructivist ideas, Posner et al.'s (1982) classical conceptual change model is based on the principle that prior knowledge and the mechanism of assimilation, accommodation, and equilibration are critical to the process of students changing their current conception to be more consistent with normative scientific knowledge (Vosniadou, 2003). The term assimilation refers to evolutionary changes that adds to and are consistent with students' existing knowledge. In contrast, accommodation pertains to revolutionary changes that results in knowledge which is distinctly different from students' prior knowledge (Dole & Sinatra, 1998). Equilibration describes the process of how change comes about in learning. Put very simply, students enter a state of disequilibrium when they are exposed to new, discrepant information. Students must then reconcile the conflicting information with their existing knowledge to attain equilibrium (Schunk, 2016). For radical conceptual change to occur, students undergo cognitive accommodation and substantially reorganize or change their current conceptions. If instruction to promote conceptual change is successful, students come to terms with discordant information and change their way of thinking to accept and correctly explain known scientific understanding (Posner et al., 1982).

The principal supposition of the prevailing conceptual change instructional paradigm is if students become aware of the conflict between their existing conception and normative scientific knowledge, then students will be compelled to change their conception (Sinatra & Pintrich, 2003a). However, this is not always the case. And the obvious question is why? Posner et al. (1982) addressed this question by stipulating there are four conditions for cognitive accommodation to take place. (1) There must be *dissatisfaction* with existing conception. (2) A new conception must be *intelligible* with its meaning clear, understandable, and applicable to the given context. (3) A new conception must appear initially *plausible* with the capacity to solve problems. (4) A new concept should suggest the possibility of *fruitful* answers and ideas, or open new areas of inquiry. The notion being if all four conditions are met, then revolutionary conceptual change is more likely. However, if none or any of these four conditions is not met, then conceptual change is less possible (Sinatra, 2005).

The underlying assumptions of the classical conceptual model have been criticized. First, students may not think and operate in the purely rational manner that is assumed. Second, students existing conceptions may not be coherent enough for them to readily discern when there is conflict between what they know and what is to be learned. Third, conceptual change for students may be closer to the evolutionary rather than the revolutionary end of the continuum. Fourth, a total emphasis on external factors such as teacher action and curriculum interventions completely disregard student characteristics that influence the process of conceptual change (Sinatra & Pintrich, 2003a). Most researchers agree students' thinking about conflicting information alone does not fully explain the process of conceptual change. A much more nuanced and complex process is at hand that better accounts for similarities in students' prior knowledge yet explains differences in conceptual change learning outcome that is frequently observed (Vosniadou, 2003). The next model discusses attempts to address criticisms of the classical model which includes adding student attributes and contextual learning factors to better explain the conceptual change process.

Extended Conceptual Change Model

The puzzling question persists as to why some students undergo conceptual change and some do not, despite having similar background knowledge. Put another way, why does prior knowledge foster restructuring of existing conceptions to resolve discrepancies in some instances, but act as a barrier in others (Sinatra, 2005)? Pintrich et al. (1993) argue that in addition to Posner et al.'s (1982) four requisite conditions for cognitive accommodation, student attributes, namely epistemic motivation, goals, personal interest, value beliefs, importance, self-efficacy, and control beliefs function as significant motivational factors that impact students' willingness to change their minds. Also, the richness of the experiences that students have in the educational context in which they learn influence whether students change their current conception. Pintrich et al.'s (1993) insistence that extrarational factors are important in the conceptual change process changed research in the field from that point forward. Pintrich et al.'s (1993) pioneering work focused attention on motivational and contextual factors that were previously ignored yet are arguably instructive to understand differences between students with similar prior knowledge but dissimilar motivations to change their conceptions. This resulted in an increasing number of researchers characterizing conceptual change as motivational, social, situational, and affective in nature, beyond its cognitive underpinnings (Sinatra, 2005). Pintrich et al.'s (1993) extended conceptual change model laid the groundwork for researchers to later understand that extrarational factors tend to impact conceptual change when there is a high level of commitment to prior knowledge, and that commitment evokes emotions that either facilitates or hinders a change in students' point of view (Sinatra & Mason, 2008). The focus on students' attributional qualities in the extended model of conceptual change contributed to a wider view of the conceptual change process found in the next model, one that stretched beyond perspectives from the domains of cognitive psychology and science education.

Cognitive Reconstruction of Knowledge Model

Dole and Sinatra (1998) developed the cognitive reconstruction of knowledge model to better explain conditions for long-lasting and strong conceptual change. They did this by building on the ideas of the extended conceptual change model (Pintrich et al., 1993) and drawing on a wealth of research from three domains, cognitive-developmental psychology (e.g. Piaget, 1972; Carey, 1992; Chi, 1992), science education (e.g. Posner et al., 1982), and social psychology (e.g. Petty & Cacioppo, 1986). Dole and Sinatra (1998) coopted the elaboration likelihood model of persuasion (Petty & Cacioppo, 1986) from social psychology. In doing so, Dole and Sinatra (1998) argue students take either of two routes towards conceptual change, depending on the message or new information they receive. One, a *central route* in which students think long and hard about the new information and what it means in relation to their current conceptions. Two, a *peripheral route* where students make a cursory assessment of the new information based on tangential cues in the environment. The central route typically leads to an enduring change in students' understanding. The peripheral route usually results in a transient change in students' conception. To convince students changing their current conception is warranted, the new information must be coherent, linking existing and new ideas into a conceptual whole. It must be rhetorically compelling; that is, intensely persuasive to the student. Finally, the new information must be comprehensible and plausible (consistent with Posner et al., 1982). If the new information lacks one or more of these characteristics, the probability of long-lasting conceptual change greatly diminishes (Dole & Sinatra, 1998).

Dole and Sinatra (1998) hypothesize the strength, coherence, and commitment of students to their existing knowledge mediate their willingness to scrutinize and change what they already know. If students' existing knowledge is coherent and detailed, then students are generally satisfied with what they know. If students' existing knowledge explains the new information satisfactorily, then students feel no need to change their current understanding. If students' existing knowledge arouses a defensive stance toward the new information, then students tend to resist accommodating the new information. Conversely, if students' existing knowledge is fragmented and ambiguous, then students are likely to be dissatisfied with what they understand. If students' existing knowledge fails to explain the new information, then students are more amenable to changing their conception. If students' existing knowledge does not provoke a distrustful attitude toward the new information, then students are more inclined to modify their thinking about what they know. The degree to which students cognitively engage with the new information ranges from high to low elaboration. If students deeply process the new information and grapple with its meaning, then this high elaboration may lead to strong conceptual change. However, if students process the new information in a superficial manner, this low elaboration often results in weak conceptual change.

Dole and Sinatra (1998) view motivation from two perspectives: (1) motivation to change current conceptions based on an *external* message, just discussed; and (2) motivation to direct *internal* actions to bring about conceptual change. For students to be motivated to undergo radical cognitive restructuring of knowledge, they must have the mental capacity to process the new information. Further, the new information must induce cognitive conflict so that students become dissatisfied with their existing conceptions (in agreement with Posner et al., 1982). Students must have a need to understand the new information (need for cognition), and they must find the new information personally relevant. Additionally, the learning environment must reinforce the value of the new information to students. All factors, persuasive message, learner characteristics, and learning environment, interact in a complex, multifaceted way to promote, if conditions are favorable, or impede, if conditions are unfavorable, the process of conceptual change. Dole and Sinatra's (1998) model marked a shift towards the intentional behavior of the learner and its impact on the process of conceptual change as seen in the next model.

Intentional Conceptual Change Model

Many researchers concede inducing cognitive conflict between existing knowledge and to-be-learned concepts, plus fostering deep cognitive engagement with new ideas are often insufficient conditions for conceptual change (Sinatra & Pintrich, 2003a, 2003b). Consequently, interest in intentional learning that places the impetus for change under the student's control grew and became more of a focus in conceptual change research. In intentional conceptual change, the student intentionally initiates and directs the interaction between internal cognitive processes and external environmental influences that bring about change (Vosniadou, 2003). Broadly speaking, intentional conceptual change dictates that learning is goal-directed, self-regulated, and purposeful, with a determined state of mind to change understanding (Pintrich & Sinatra, 2003;

Sinatra & Pintrich, 2003a). Student attributes that serve to augment intentional level of awareness that conceptual change is necessary include, but are not limited to, personal interest, self-efficacy, control beliefs, and achievement goals (Sinatra & Mason, 2008). There is no normative intentional conceptual change model. Instead, there are multiple perspectives on what intentional conceptual change is and under what circumstances it does and does not occur. More work needs to be done to develop a definition of intentional conceptual change that is widely accepted by the research community. Also, researchers need to specify what structures and processes are involved and what role contextual factors play (Pintrich & Sinatra, 2003). Despite the obvious challenges for this body of work, there are notable contributions in intentional conceptual change research that can help clarify individual processes that operate during cognitive restructuring of knowledge. One such example is Cynthia Hynd's (2003) discussion on intentional conceptual change in response to persuasive messages.

Hynd (2003) characterizes intentional conceptual change as a motivated metacognitive response to persuasive information that is accompanied by an intentional change in understanding. Building on Dole and Sinatra's (1998) position that there are two routes to persuasion, Hynd (2003) argues students undertake *both* central and peripheral processing of new, persuasive information. Students may use one or the other route under different circumstances; or they may use both routes at the same time, depending on the situation. Hynd (2003) reasons students use central processing to evaluate the strength of new information and how it integrates into a believable, consistent conceptual whole. Whereas, students use peripheral processing when they focus on the new information's appeal, its source, and the context in which it is received. When students' prior knowledge is correct or more complete, they are more likely to understand new information and use central processing. In a similar vein, students tend to undertake central processing when new information is moderately different from their existing knowledge; that is, not too close to be ignored and not too distant to be dismissed. On the other hand, when students have no or incomplete prior knowledge, they are more likely to rely on peripheral cues to process new information. Hynd (2003) asserts the use of both central and peripheral routes may be superior to using central processing alone. Hynd (2003) cautions that although conceptual change is possible with just peripheral processing, this type of conceptual change is typically weak and less likely to occur if students are not motivated, the discrepant information has no personal relevance, and the learner has no need for cognition regarding the new information.

Analysis of Conceptual Change Models

Theories and models are continually refined as more information becomes available that improves our description, explanation, and basic understanding of a phenomenon. In the case of conceptual change, subsequent theoretical models elaborate on previous ones by providing useful insights into aspects of the nature of learning generally, and conceptual change learning especially. It is therefore important that conceptual change researchers avoid replacing one aspect of conceptual change from an earlier model for another aspect introduced in a later model. For instance, conceptual change researchers should not deemphasize rational, cognitive constructs to capture extrarational attributional, affective, and contextual constructs, or the other way around. If they do, conceptual change researchers run the risk of limiting their understanding of a highly complex process. Sinatra and Mason (2008) advocate for a nuanced model of conceptual change, one that accounts for complementary facets comprised of cognitive, attributional, affective, and contextual components which provide critical yet very different aspects of a complex process. This multifaceted view of conceptual change offers a richer understanding of the process and affords a fuller description for operationalizing a model of conceptual change.

Operational Model of Conceptual Change

Theoretical models of conceptual change must endeavor to describe and explain the complex, multifaceted nature of the conceptual change process. According to Vosniadou (2003), a full theoretical model of conceptual change should provide information about four important facets: (1) individual cognitive changes; (2) effects of individual motivational and affective factors; (3) the educational setting in which teaching, and learning take place; and (4) the broader social and cultural environments in which students live and learn. In other words, a comprehensive model of conceptual change must account for a rational and extrarational view of learning. Under ideal circumstances, an operational model of conceptual change would include all four facets that Vosniadou (2003) mentions. However, practical constraints may make this difficult. Even so, at the very least, an operational model of conceptual change should include cognitive, motivational, and contextual constructs, as Pintrich et al. (1993) propose. In this study, an operational model of conceptual change is presented with conceptual change construct of evolution by natural selection, cognitive construct of formal reasoning ability, attributional construct of academic motivation, and contextual construct of teaching practice. Operationalizing a model of conceptual change necessitates using psychometrically vetted instruments that correctly and consistently measure the variables under study. To this end, in the following subsections, I trace the development of instruments designed to measure conceptual understanding of evolution by natural selection, formal reasoning ability, academic motivation, and teaching practice, respectively. For each construct, I take a historical stance to discuss seminal works involved in the inception and refinement of the instrument for wide-scale use. I accentuate the psychometric methods used to establish validity and reliability. Also, I highlight any trends to streamline the overall effectiveness and ease in administering and scoring the instruments.

Conceptual Change Construct of Evolution by Natural Selection

Understanding key ideas in the theory of evolution has proven to be difficult for many students. Not surprisingly, researchers have found it challenging to design instruments that properly assess change in students' conceptual understanding of evolutionary biology. Three principal objectives have shaped instrument development to test students' understanding of evolution by natural selection. First, an instrument is desired that correctly measures students' knowledge, is appropriate for a pre-test/post-test format, easily administered, and quickly scored. Second, a credible instrument must have sound psychometric properties that show clear evidence it is a valid and reliable measure of students' knowledge of evolution by natural selection. Third, an effective instrument must discriminate among students' understanding of evolution by natural selection that is of high and low ability. A discussion follows that addresses each of these three issues in the development of instruments to measure conceptual change of evolution by natural selection.

One of the most important objectives in developing an instrument that assesses change in students' conceptual understanding of evolution by natural selection is to cataloge students' correct conceptions and misconceptions, plus measure the effects of instruction on students' understanding. To accomplish this feat, Bishop and Anderson (1990) initially introduced the open-ended essay test, which was afterward revised by Nehm and his colleagues and presented as the Open-Response Instrument (*ORI*, Nehm & Reilly, 2007; Nehm & Schonfeld, 2008). Bishop and Anderson concentrated on biology nonmajors, while Nehm and his colleagues focused on biology majors of diverse ethnic backgrounds in the developing their respective instrument. This gave credence to the aptness of the open-ended test format for assessing the knowledge of all introductory biology students, whether they were majors or non-majors. Twelve years after Bishop and Anderson (1990) published the very first open-ended essay test instrument, Anderson et al. (2002) created a closed-ended multiple choice test instrument, the Conceptual Inventory of Natural Selection (*CINS*), aimed at producing a more efficient test for measuring students' understanding of evolution by natural selection. Nehm and Schonfeld (2008) carried out in-depth comparisons of *CINS* and *ORI* using older classical test theory (CTT) analyses and more recently developed item response theory (IRT) analyses. They concluded both the *CINS* and *ORI* instruments are (*a*) suitable for testing biology majors and nonmajors, (*b*) produce similar measures of key ideas central to the concept of evolution by natural selection, and (*c*) could serve as replacements for labor-intensive oral interviews. However, Nehm and Schonfeld (2008) pointed out *ORI* is better than *CINS* at capturing a wider range of student misconceptions, and new questions were needed for the *CINS* to adequately discriminate among high performing students. Nevertheless, Nehm and Schonfeld (2008) conceded the *CINS* is more expeditious than *ORI* at assessing students' understanding of the concept of evolution by natural selection.

The closed-ended multiple choice test format had great appeal because of its efficacy in evaluating change in students' understanding of the concept of evolution by natural selection for large classes. Following Anderson et al.'s (2002) example with *CINS*, Kalinowski et al. (2016) developed the Conceptual Assessment of Natural Selection (*CANS*), published 14 years later, with the intention of using more sophisticated psychometric methods to improve instrument validity and reliability. Anderson et al. (2002) used traditional methods of principal component analysis (PCA) and CTT analytical procedures to develop *CINS*, whereas Kalinowski et al. (2016) took advantage of more powerful IRT techniques to develop *CANS*. Moreover, Kalinowski et al. (2016) benefited from meeting updated educational and psychological testing standards (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014), which increased the rigor of the instrument development process. Notwithstanding, there are important similarities and differences between the instruments. With regards to face validity, both *CINS* and *CANS* attempt to test key ideas central to the concept of evolution by natural selection. *CINS* was designed to test 10 key ideas (biotic potential, population stability, natural resource is limited, limited survival, variation within a population, variation is inherited, differential survival, change in population, origin of species, and origin of variation), with 2 questions each. In contrast, *CANS* was designed to test 5 key ideas (variation, selection, inheritance, mutation, and the interaction of these four to bring about evolution), with 3 questions on variation, 5 questions on selection, 4 questions each on inheritance and mutation, and 8 questions on evolution. Increasing the number of questions for each key idea enhances how well the instrument measures students' understanding of each idea. Both *CINS* and *CANS* are structured similarly with each instrument divided into sections that ask questions about a different species. For each section, information relating to the species is followed by several questions with relatively short prompts. All questions are formatted in the same way with response item choices having one correct answer and two or more distractors that address common misconceptions about evolution by natural selection. The *CINS* is a 20-item test that asks questions about Galapagos finches, Venezuelan guppies, and Canary Island lizards. Whereas, the *CANS* is a 24-item test that asks questions about anteaters, bowhead whales, saguaro cacti, and mosquitoes. The questions on the *CANS* are about three animals and one plant, while the questions on the *CINS* are only about animals. Kalinowski et al. (2016) made a concerted effort to vary the context of the questions as students may think about questions with dissimilar contexts differently. For instance, students may think about a trait associated with an animal differently from how they think about the same trait associated with a plant. Or, students may think about trait loss differently from trait gain. Purposely, the questions on the *CANS* were designed to exhibit greater variability in context than questions on the *CINS*.

To determine the effects of instruction on change in students' understanding of the concept of evolution by natural selection and to establish construct validity, a pre-test/post-test design is used for both *CINS* and *CANS*. In the final stages of instrument development, *CINS* was administered as a pre-test to 206 students (split into two sections of 110 and 96 students). PCA with varimax rotated component patterns indicated there were 7 components (not 10 consistent with the intended key ideas for the instrument) that account for 53% of the total variance for the 20 items, with factor loadings of >0.40 on at least 1 component. Only 1 item had a factor loading of >0.40 on multiple components (Anderson et al., 2002). In comparison, *CANS* was administered as a pre-test to 233 students and a post-test to 266 students in the same cohort. Only 218 students completed both the pre- and post-tests for the cohort. Students, on average, scored higher on the post-test compared to the pre-test. IRT analyses of data obtained from 218 students on the pre-test (those who also completed the post-test) suggest 18 of 24 questions measure the same latent construct, presumably ability to understand evolution by natural selection. All 8 evolution questions, all 4 mutation questions, all 4 inheritance questions, and 2 of 5 selection questions (but none of the variation questions) loaded on the single latent construct with factor loadings >0.65 (Kalinowski et al., 2016). Two noteworthy advantages of *CANS* over *CINS* are (*a*) *CANS* measure a discernable increase in conceptual change for the post-test over the pre-test, and (*b*) *CANS* convincingly measure a single construct of students' ability to understand the concept of evolution by natural selection.

Customarily, the precision of an instrument has been used as a benchmark to evaluate its quality and usefulness. In the CTT tradition, reliability is an indication of an instrument's precision, or the degree to which the true score and observed score correlate for the items. Specifically, reliability is the quotient of variance of true score divided by the variance of observed score, and standard error is the difference between the true score and the observed score. The computation of CTT reliability coefficients, such as Cronbach's alpha (α) and Kuder-Richardson 20 (KR_{20}), assume the standard error for the items is constant (Adams, 2005). For IRT analytical techniques, an instrument's precision is instead related to the item information for the latent trait being measured by the instrument. Generally, as item information for the latent trait increases, the standard error for the item decreases (Chalmers, 2012). Reliability is not viewed as an ideal method for determining the precision of an instrument in IRT analyses (Adams, 2005; Chalmers, 2015a). This is because IRT analyses generate item, scale, or test information functions that provide latent trait estimates and their associated standard errors. It is therefore considered a step backward to reduce a complex conditional function to a single aggregate number (Chalmers, 2015b). Nevertheless, the preeminence of reliability coefficients abounds. But unfortunately, many practitioners do not have a clear understanding of under what circumstances and with which statistical analyses reliability coefficients are meaningful (Adams, 2005). It seems that because of the prominence of reliability coefficients, some IRT programs include arguments to determine the overall reliability, although its use is discouraged because there are multiple methods to calculate its value and many practitioners are unsure which method to use and how to choose. Besides that, estimated reliability coefficients are redundant within the context of IRT (Chalmers, 2012, 2015a, 2015b). Despite the inherent contradiction, one method used to estimate overall reliability is empirical reliability in mirt: A multidimensional item response theory package for the R environment. The empirical reliability for the instrument is computed as the estimate of the true latent trait for all items divided by the sum of the estimate of the true latent trait for all items and the estimate of the associated standard errors for the latent trait (Chalmers, 2012, 2018). Comparing the precision of *CINS* and *CANS* in terms of reliability coefficients, KR_{20} values of 0.58 (110 students) and 0.64 (96 students) were reported for *CINS*, with 0.60 or higher denoted as reasonable (Anderson et al., 2002). While, empirical reliability values of 0.88 for the pre-test, and 0.87 for the post-test were reported for *CANS*. The same sample of students yielded α values of 0.85 and 0.86 for the pre-test and post-test, respectively, for *CANS* (Kalinowski et al., 2016). Cronbach's alpha values of $0.8 \leq \alpha < 0.9$ is commonly considered good. These reliability coefficients suggest both *CINS* and *CANS* meet the acceptable threshold for internal consistency; however, *CANS* seem to be a much more precise instrument than *CINS* at measuring change in students' understanding of the concept of evolution by natural selection.

In furthering the evaluation of *CINS* and *CANS*, a principal goal of diagnostic instruments that assess students' understanding of the concept of evolution by natural selection is to distinguish between students who demonstrate mastery of the concept versus those who do not. Two properties, item difficulty and item discrimination, give some indication of the variability in students' performance on the instrument, and therefore a measure of the range of students' conceptual understanding of evolution by natural selection. Item difficulty, in the traditional sense, is defined as the percentage of students who selected the correct response. Also, item discriminability is often computed using the point biserial value method, where a value close to 1.00 means greater discriminability and below 0.30 is considered poor. Item difficulty on *CINS* ranged from 14.5% to 80.6% with a test average of 46.4%, close to the recommended 50% test difficulty for a typical class. Of the 20 items on the *CINS*, 6 were below the 0.30 minimum for item discriminability, and only 1 item was above 0.50 (Anderson et al., 2002). In contrast to conventional methods, IRT analytical procedures for evaluating item difficulty and item discrimination involve the interpretation item trace lines. For the *CANS*, item difficulty, defined as the logit of student ability when the probability of a correct response is 50%, ranges from approximately −1.25 to +1.25 standard deviations from the mean, with the mean centered at 0. Arguably, 21 of 24 questions have a reasonable range of difficulty for the key ideas being tested. That said, only 2 of 5 selection questions work well for item difficulty. The steepness of the item trace line is a measure of how well the item discriminates among student ability. The steeper the item trace line, the greater the item discrimination; that is, the better the item discriminates between the probability of a correct response for students whose ability differ by only small amounts. The item discrimination for 19 of 24 questions is very good for *CANS*

(Kalinowski et al., 2016). For sure, the sophistication of the IRT analytical methods give strong support for why *CANS* is a better instrument than *CINS*.

Even more evidence about the performance of *CANS* is gleaned from the item difficulty, item discrimination, and the guessing rate per item combined that provide IRT information for each item, or the scale if summed over all items. The item or scale IRT information is represented graphically as a function of the logit of student ability, referred to as the information curve. For the *CANS*, IRT information curves for all 8 evolution, all 4 mutation, all 4 inheritance questions, and 2 of 5 selection questions show high information to explain students' ability to understand evolution by natural selection for a reasonable range of student abilities. The *CANS* test information curve suggests the instrument seem to do a better job of estimating student ability to understand evolution by natural selection for high-ability students than it does for low-ability students (Kalinowski et al., 2016). In summary, *CANS* is demonstrably valid and reliable; moreover, *CANS* have advantageous psychometric properties that differentiates among students' ability to understand the concept of evolution by natural selection. This makes *CANS* particularly fitting to measure the effects of instruction on conceptual change for this study.

Cognitive Construct of Formal Reasoning Ability

There is credible evidence many individuals develop certain reasoning patterns during adolescence to adulthood that play an important role in their ability to carry out scientific investigations and learn scientific concepts (Lawson, 2004; Lawson et al., 2000). Some researchers have attributed this type of abstract reasoning to students' formal reasoning ability, while others have characterized it more broadly as students' scientific reasoning ability (Kalinowski et al., 2018). Rooted in Piagetian cognitive theory are hypothetico-deductive reasoning structures described as formal operations that enables an individual to reason about verbally stated hypotheses and deeply intellectual, yet intangible ideas. These reasoning structures decisively extend beyond mere manipulation of concrete objects (Piaget, 1972). They are logical reasoning processes that guide the search for and the evaluation of evidence to support or reject hypothetical causal propositions required in scientific inquires (Lawson, 1978). One of the most profound contributions to our understanding of formal operational thinking was Piaget's genius to design intellectual tasks that externalized an individual's thinking into observable behavior. This allowed trained cognitive psychologists to draw inferences about the underlying mental processes and logical reasoning patterns operating based on an individual's execution of these intellectual tasks and the explanations they give while performing these tasks (Bruner, 1959).

Inherent to Piagetian cognitive theory is the idea that students' reasoning as they carry out select intellectual tasks is related to certain logical operations common to all the tasks. For instance, Piagetian tasks that require formal operational thinking include (*a*) isolation and control of variables, (*b*) combinatorial reasoning, combinatorial analyses of all possible causes, (*c*) proportional reasoning, determining the proportional relationship or constant ratio among variables, (*d*) probabilistic reasoning, recognizing the probabilistic nature of phenomena, and (*e*) correlational reasoning, weighing of evidence to confirm or disconfirm if variables are related (Lawson, 1978; Piaget 1964). Piagetian tasks have taken on great significance in the field of science education, and over time, they have been modified and used widely as diagnostic tools. In this section, I discuss significant changes in practice to test students' formal reasoning ability, which largely pivoted from the observation of students performing Piagetian tasks in clinical interview settings to the general dissemination of paper-and-pencil instruments in science classrooms. Tracking this history is helpful in advancing our understanding of three critical features important to the development of instruments that test students' formal reasoning ability: (1) what do the instruments attempt to measure, (2) how well do the instruments measure what they set out to do, and (3) how suitable are the instruments for investigating relationships among formal reasoning ability, science learning, and conceptual change.

Many of the earlier instruments developed to test students' formal reasoning ability heavily resembled the Piagetian clinical interview of an observer evaluating an individual's performance on intellectual tasks. One such instrument by Sayre and Ball (1975) was constructed from 5 modified Piagetian formal operational tasks that tested hypothetico-deductive reasoning, control of variables, combinatorial reasoning, and proportional reasoning. A specially trained interviewer administered the instrument to one student at a time. The interviewer classified each student as either formal operational or nonformal operational consistent with the student's answers to probing questions and their performance on each task. Students' overall assessment was based on all 5 tasks. Another instrument developed by Shayer, Adey, and Wylam (1981) comprised 7

Piagetian tasks ranging from preoperational to late formal operational. In this case, an administrator demonstrated each task for a group of approximately 30 students over a 40 to 50-minute period. Students recorded their open-ended response to questions on each task, which were subsequently scored and tallied using an agreed upon rubric. Students were classified into developmental levels according to their total score. These are two examples of earlier instruments that required trained personnel to administer and appraise students' performance on Piagetian tasks. There was obvious improvement in moving from individual to group testing, and certainly this was a step in the right direction. The problem was, however, running and scoring these instruments was still time consuming and labor-intensive for all persons involved in the process.

Science educators became increasingly interested in creating valid and reliable paper-and-pencil instruments that retained positive aspects of the Piagetian clinical interview yet had the advantage of assessing groups of students in a relatively short time. Lawson (1978) made considerably strides in developing an instrument with closed-ended test items that improved the process of administering the instrument and greatly reduced the time required for scoring and evaluating students' formal reasoning ability. Lawson's (1978) 15-item instrument consisted of concrete and formal operational Piagetian tasks. Of the 15 tasks on the instrument, 12 were previously published (on conservation of mass, displaced volume, isolation and control of variables, proportional reasoning, as well as combinatorial reasoning), and 3 on probabilistic reasoning were developed by Lawson (1978). Each task was demonstrated for students, and then students responded to test items that asked them to check the box with the correct response regarding the task and explain why they chose that answer. Items were scored as a point if the correct answer was chosen and an adequate explanation was given for choosing the answer; otherwise no point. Lawson (1978) used three types of evidence to validate his instrument. (1) Face validity, for which a panel of experts determined the test items included a reasonable selection of Piagetian tasks. (2) Concurrent validity, that Lawson refers to as convergent validity (Roberts, 1980), and others refer to as criterion-related validity (Tobin & Capie, 1981), which is the correlation between students' total test score on the instrument and summed scores on select Piagetian interview tasks. Lawson (1978) reported a correlation of 0.84 between student scores on the instrument and Piagetian tasks. Finally, (3) construct validity based on PCA with varimax rotation of test items on the instrument, as well as select Piagetian concrete and formal operational tasks. Lawson (1978) found 3 principal components (as opposed to 2 with 1 each for concrete and formal reasoning) that explained 66.0% of the total variance for the test items. Also, the select formal operational Piagetian interview tasks loaded moderately on 2 factors (not heavily on 1 as expected). For estimate of reliability, Lawson (1978) used KR_{20} to assess the internal consistency of the instrument and determined a value of 0.78. These evidences provide encouraging, but not resounding support for the validity and reliability of the instrument. By Lawson's (1978) own admission, the results for construct validity are tentative at best since they only represent analyses for 72 students who completed the test instrument the select Piagetian interview tasks.

Researchers continued to work toward developing an instrument that efficiently and accurately assessed students' formal reasoning ability. One example is Tobin and

Capie's (1981) Test of Logical Thinking (*TOLT*) constructed using 10 select tasks that explicitly test formal operational reasoning, all taken from Lawson's work (Lawson, 1978; Lawson, Adi, & Karplus, 1979). The tasks were control of variables, proportional reasoning, combinatorial reasoning, probabilistic reasoning, and correlational reasoning. *TOLT* improved on the structure and format of Lawson's instrument in three ways. First, they used a video-tape to describe the context for each question, instead of having a specially trained demonstrator do it. Second, the test items consisted of graphical illustrations that elaborated on the context of the problem where necessary, which again negated the need for a trained demonstrator to provide layers of related detail. Third, they used a multiple-choice question format for both the items that tested students' reasoning and the follow-up item that asked students to justify their chosen answer. This removed the need for specially trained persons to do the scoring, and it also meant scoring was done faster. In validating *TOLT*, Tobin and Capie (1981) included predictive validity as an added source of evidence and reported correlations between students' scores on their instrument and established standardized tests such as the Scholastic Aptitude Test (SAT®). Tobin and Capie's (1981) remarkable achievements in developing their instrument seemingly had an impact on Lawson's subsequent work.

There was a noticeable shift in Lawson's work away from demonstrations to provide context for tasks towards graphics and detailed descriptions in self-contained questions. For instance, a proportional reasoning question that was previously demonstrated with measuring cylinders of the same height and different diameters became a question with a stick figure illustration and vivid text that described how stick figures of different heights are measured using paper clips of the same length (Johnson & Lawson, 1998; Lawson, 1978). Lawson's original instrument did not include questions on correlational reasoning. However, with the assistance of his colleagues, Lawson developed two correlational reasoning questions: one with a correlational relationship related to mice and one without a correlational relationship associated with fish (Lawson et al., 1979). Also, the original instrument did not have questions that explicitly test hypotheses with tentative causal agents that were abstract entities and unobservable to the naked eye (e.g. invisible gases or ions in a solution). Lawson and his colleagues developed two questions about a burning candle and red blood cells for this purpose (Lawson et al., 2000). All new questions developed after the publication of the original instrument were complete with diagrams and descriptive text to fully explain the context of the problem and what was being asked of students. Lawson (2000a) modified his original instrument to incorporate new questions that test formal operational reasoning skills not addressed in the previous version. All questions, including those that ask students to justify their selected answer, now had a multiple-choice question format. This modified version, the Classroom Test of Scientific Reasoning (*CTSR*), consisted of two types of concrete operational reasoning (conservation of mass and displaced volume) and five types of formal operational reasoning (control of variables, proportional, probabilistic, correlational, and hypothesis testing with causal agents that are unobservable). Even though *CTSR* is arguably quite different from the instrument originally published, Lawson (2000a) did not independently provide any evidence for the modified instrument's validity and reliability.

Concerned about the veracity of *CTSR* to accurately and consistently assess students' formal reasoning ability, Kalinowski and his colleagues carried out IRT analyses to ascertain whether *CTSR* was valid and reliable, or if a new instrument was needed (Kalinowski et al., 2018). Analyses of data obtained from several introductory science courses indicated students responded to certain pairs of questions on *CTSR* similarly. Students did not seem to discern any difference about what was being asked when identical scenarios were used to provide context for the same type of formal reasoning questions. For example, students' response to two separate control of variable questions with similar contexts were highly correlated. Although the causal agent for the illustrated effect for each control of variable question differed, both questions were related to fruit flies in tubes exposed to light. Students seems to either assume or reason that there was no difference in the causal agent for each case because both questions were about fruit flies. Kalinowski et al. (2018) point out this type of item dependency among questions inflates estimates of reliability and increases the apparent discrimination of questions, which they contend must be fixed. Kalinowski et al. (2018) initially considered correcting the issue of item dependency with *CTSR*, but eventually decided it was better to construct a new instrument. The development of Kalinowski et al.'s (2018) instrument, the Montana State University Formal Reasoning Test (*MSU-FORT*), involved 22 drafts administered to about 7500 students in introductory science courses with 32 rounds of testing. The resulting 20-item *MSU-FORT* instrument comprise 5 types of formal reasoning (control of variables, proportional, probabilistic, correlational, and hypothesis testing) with 4 items each. The *MSU-FORT* incorporate 5 items (3 unmodified and 2 modified) from *CTSR*. IRT analyses of data obtained from 236 students in an introductory biology course suggest *MSU-FORT* measures a single latent factor, expectedly students' formal reasoning ability. Item difficulty interpreted from the item trace lines is quite good for 14 of 20 questions. The steepness of the item trace lines indicates very good discrimination for 5 and reasonable discrimination for another 5 of 20 questions. There are no obvious problems with item dependency for pairs of questions testing the same type of formal reasoning. The reported empirical reliability for *MSU-FORT* is 0.81. It is also worth noting there is reasonably good correlation (0.64) between *MSU-FORT* and *CTSR* for 172 students that completed both instruments. Given the more rigorous psychometric properties of *MSU-FORT* compared to *CTSR*, *MSU-FORT* is convincingly a better measure of students' formal reasoning ability.

The question of whether formal reasoning is specifically tied to subject-matter knowledge or more broadly related to integrated science process skills became a focus of inquiry because of its implications for science learning; bringing attention to the question of which instruments are most suited to measure formal reasoning within domain-general contexts. In tackling this issue, Tobin and Capie (1982) argue generic scientific reasoning skills are needed to conduct experiments for quantitative investigations. Tobin and Capie (1982) identified 12 steps that are necessary to successfully carry out an investigation. For instance, how to determine the respective dependent and independent variable(s), and what procedures are best suited to manipulate the independent variable(s) to achieve the stated objective. Tobin and Capie (1982) developed 2 questions for each step, which resulted in a 24-item multiple-choice instrument, the Test of Integrated

Science Processes (*TISP*), they reason is appropriate for testing students' integrated science process skills. Lawson (1992) addressed the matter of content knowledge free formal reasoning from a different angle and proposed the multiple-hypothesis theory of advanced scientific reasoning. Lawson (1992) argue there is some evidence advanced scientific reasoning is the ability to test hypotheses in a manner that moves thinking beyond one cause for one effect to multiple causes for specific effects. Furthermore, Lawson (1992) argue advanced scientific reasoning is the general disposition to consider alternative possibilities and to search for alternatives, even when those alternatives are not immediately apparent. Acknowledging the possibility that formal reasoning may not be entirely divorced from subject-matter knowledge, Lawson (1992) reason the context of the problem probably influences a student's ability to imagine an alternative cause for a given effect. This likely informs the conclusion(s) a student eventually draws. This begs the question: which type of instrument is most fitting to test students' ability to learn science and change misconceptions about science? Is it the type of instrument, for example *TISP*, that tests broad-based integrated science process skills to carry out investigations? Or, is it the type of instrument, for example *MSU-FORT*, that vigorously assesses how well students test hypotheses and use formal operational reasoning to find causal-effect relationships? I contend *MSU-FORT* is a much more thorough measure of (*a*) students' scientific reasoning skills to plan and carry out investigations, (*b*) ability to correct misconceptions about scientific concepts, as well as (*c*) students' disposition to find plausible solutions for interesting questions in a variety of contexts. Therefore, to investigate the effects of formal reasoning ability on conceptual change for this study, *MSU-FORT* is the best option.

Attributional Construct of Academic Motivation

Motivation to learn may be defined as students' predisposition to find academic pursuits meaningful and worthwhile; that is, a habit of mind symptomatic of a tendency to derive tangible benefits from learning activities (Glynn et al., 2007). In keeping with this definition, motivation to learn cannot be observed directly. Rather, it can only be determined indirectly. The cognitive model of motivation that identifies students' perception of themselves and the task at hand as the most important components of motivation has a long tradition in psychological research (McKeachie, Pintrich, Lin, & Smith, 1986). This model utilizes intrinsic goal orientation, extrinsic goal orientation, self-efficacy, self-determination, task value, test anxiety, and amotivation as scales that tap into and quantify students' conscious understanding of their psychological state to be motivated to carry out an action (e.g. Glynn & Koballa, 2006; Midgley et al., 2000; Pintrich, Smith, Garcia, & McKeachie, 1991; Vallerand et al., 1992). The most commonly used method for measuring motivation is individual self-report on various motivation scales. The obvious shortcoming with this approach is self-report may be biased or incomplete, and it may not fully or accurately represent the psychological state that facilitates an individual's action (Touré-Tillery & Fishbach, 2014). Despite these concerns, self-report questionnaires are widely used because they are more efficient, less time consuming, and more cost-effective than other options such as evaluation of interviewees in one-on-one interviews and making inferences about individuals' motives based on observed behavior (McKeachie et al., 1986).

The question of what motivates students to learn science has been a subject of investigation for many years. Glynn and Koballa (2006) identified six constructs, intrinsic motivation, extrinsic motivation, personal goals, self-efficacy, self-determination, and assessment anxiety, that they hypothesize influence students' motivation to learn in college science courses. They developed the Science Motivation Questionnaire (*SMQ*) that consists of six scales, with five items each, corresponding to the six motivational constructs mentioned above. Using a 5-point Likert scale ranging from 1 (*never*) to 5 (*always*), Glynn and Koballa (2006) crafted items such as (*a*) "I find learning the science interesting," for intrinsic motivation, (*b*) "Earning a good science grade is important to me," for extrinsic motivation, (*c*) "I think about how the science I learn will be helpful to me," for personal goals, (*d*) "I am confident I will do well on the science labs and projects," for self-efficacy, (*e*) "I use strategies that ensure I learn science well," for self- determination, and (*f*) "I worry about failing the science tests" for assessment anxiety (pp. 29-31). But unfortunately, Glynn and Koballa (2006) did not report any validity evidence for the instrument. They only reported the Cronbach's alpha (α) reliability coefficient for the entire instrument (α=0.93), not for the individual motivation scales as would be expected.

Over the next several years, Glynn and his colleagues presented evidence to validate the *SMQ* in a sequence of publications. Three years after the *SMQ* was published, Glynn et al. (2009) conducted PCA that indicated there were five factors, not six, that explained 60.23% of the variance. Based on the item loadings from the analysis, the five factors were labeled (1) intrinsic motivation and personal relevance, (2) self-efficacy and assessment anxiety, (3) self-determination, (4) career motivation, and (5) grade motivation. Glynn et al. (2009) noted, of the variance explained, 30.09% was attributed to intrinsic motivation and personal relevance, 13.4% to self-efficacy and assessment anxiety, 8.59% to self-determination, and 8.14% to extrinsic motivation (grade, 3.49% and career, 4.65%). The corresponding Cronbach's alphas (α) were 0.91, 0.88, 0.74, 0.88, and 0.51 for factors 1-5 listed above. Another two years passed before Glynn and his colleagues published a modified instrument, the Science Motivation Questionnaire II (*SMQ-II*, Glynn et al., 2011). Adopting the five-factor structure from the previous version, Glynn et al. (2011) developed five scales, namely, intrinsic motivation, self-efficacy, self-determination, grade motivation, and career motivation, with five items each. They carried out PCA and confirmatory factor analysis (CFA) on two separate samples to validate the modified instrument (Glynn et al., 2011). The PCA reaffirmed there are five factors that now accounted for 67.64% of the variance, with factor loadings of >0.40 for each of the five factors. Five independent measures of fitness indicate CFA yielded a good fit for a five-factor model. In addition, factor loadings of >0.40 on each factor further supported a five-factor model or five latent motivational constructs, each with five items. The reliability coefficients estimated for α are intrinsic motivation (0.89), self-efficacy (0.83), self-determination (0.88), grade motivation (0.81), and career motivation (0.92). The combined evidence for validity and reliability gives good support for *SMQ-II* as a trustworthy and accurate assessment of students' motivation to learn science. Therefore, the *SMQ-II* is an appropriate instrument for measuring students' academic motivation for this study.

Contextual Construct of Teaching Practice

It is widely acknowledged the practice of teaching is multi-dimensional, difficult to define, and tough to neatly categorize. Despite this, researchers have actively sought after some agreed upon characteristics of what is "good quality" teaching and how best to evaluate its merit. In tackling this issue, Ramsden (1991) takes the position that although ascertaining what is good quality teaching at the postsecondary level is complicated, empirical studies point to aspects of teaching that are positively related to desirable student learning outcomes. Ramsden (1991) goes on to make the point that postsecondary students, as "consumers" of a cross section of teaching practices throughout their academic careers, are credible at and capable of assessing teaching they perceive helps them learn. Therefore, students' judgment about the quality of teaching and the learning environment matters and should be considered seriously. Notwithstanding, the proposition that teachers' perspective of the teaching and learning context is valuable and vital to any discussion about improving student learning cannot be overlooked. Biggs (1989) argues genuine improvements in student learning involve an *interactive approach* to teaching in which teachers assume the quality of the learning outcome depends on how students construct their knowledge and their level of cognitive engagement in the learning tasks. Accordingly, teachers need to prioritize organizing their instruction to maximize students' deep, rather than, surface approach to learning.

Also, good quality teaching includes a *contextual approach* to teaching in which teachers become adept at discerning what learning strategies students employ and using this knowledge to teach in a manner that guides students towards methods that favor a deep approach to learning. This necessitates that teachers assess student learning outcomes in a way that enables them to determine what student misconceptions persist, so that they can appropriately adjust their teaching to better facilitate conceptual change (Biggs, 1989, 1993).

Colleagues Keith Trigwell and Michael Prosser spearheaded a series of investigations that helped illuminate our understanding of postsecondary science teachers' perspective on their approach to teaching. Prosser, Trigwell, and Taylor (1994) carried out a qualitative study on 24 science teachers who teach first-year courses at two universities and found that teachers use only a few distinct descriptors to articulate their conception of learning and teaching, respectively. Teachers described learning in primarily five ways as accumulating more information, acquiring concepts to satisfy external demands, acquiring concepts to satisfy internal demands, conceptual development to satisfy internal demands, and conceptual change to satisfy internal demands. Interestingly, teachers characterized their view of teaching in basically six ways as transmitting concepts of the syllabus, transmitting teachers' knowledge, helping students acquire concepts of the syllabus, helping students acquire teacher knowledge, helping students develop conceptions, and helping students change conceptions. Curious about the relationship between postsecondary science teachers' intentions for their teaching and the instructional strategies they reportedly used, Trigwell, Prosser, and

Taylor's (1994) probe revealed a combination of four intentions and three strategies that yielded five different approaches to teaching. These approaches are (1) a teacher-focused strategy with the intention of transmitting information, (2) a teacher-focused strategy with the intention that students acquire the concepts of the discipline, (3) a teacher-student interaction strategy with the intention that students acquire the discipline, (4) a student-focused strategy aimed at students developing their conception, and (5) a student-focused strategy aimed at students changing their conceptions. Trigwell and Prosser (1996a) concluded the relationship among conception of teaching, conception of learning, and approaches to teaching may be summarized, in general terms, as teachers who conceive of learning as information accumulation to meet external demands usually conceive of teaching as transmitting information and tend to use teacher-focused instructional strategies. In contrast, teachers who conceive of learning as developing and changing students' conceptions typically conceive of teaching as helping students to develop and change their conceptions and often engage in student-focused instructional strategies in their teaching.

Capitalizing on knowledge gained from investigating postsecondary science teachers using qualitative methods, Trigwell and Prosser turned to quantitative methods to develop an instrument aimed at capturing teachers' intentions and strategies used in their teaching. This gave birth to the Approaches to Teaching Inventory (*ATI*, Trigwell & Prosser, 1996b) that began as a 49-item instrument with five scales; namely, information transmission intention, conceptual change intention, teacher-focused strategy, student-teacher interaction strategy, and student-focused strategy. Subsequent modifications following revisions based on item reliability analysis and PCA with varimax rotation resulted in a 22-item instrument with only two scales; specifically, information transfer intention and teacher-focused strategy (*ITTF*) in addition to conceptual change intention and student-focused strategy (*CCSF*). Sample items on the *ITTF* scale are "I feel it is important to present a lot of facts in the classes so that students know what they have to learn for this subject" for the intention portion, and "I design my teaching in this subject with the assumption that most of the students have very little useful knowledge of the topics to be covered" for the strategy element (Trigwell, Prosser, & Waterhouse, 1999, p. 63). Corresponding items for the intention and strategy components on the *CCSF* scale are "I feel a lot of teaching time in this subject should be used to question students' ideas," and "We take time out in class so that students can discuss among themselves the difficulties that they encounter studying this subject" (Trigwell et al., 1999, p. 63). Further modifications to improve the construct validity of the *ATI* yielded a 16-item instrument with *ITTF* and *CCSF* scales, each consisting of 8 items (Trigwell & Prosser, 2004; Prosser & Trigwell, 2006). While Trigwell and Prosser's work provide insight into approaches to teaching for postsecondary science teachers, their work does not presume to represent the breadth of teaching practices in introductory science courses for a wide range of higher education institutions.

Since the early 2000s, increased interest in developing a framework to articulate teaching practices in use at postsecondary institutions has resulted in various categorizations. Dancy and Henderson (2007), in their study, classify teaching practices in terms of practices consistent with traditional instruction, defined as minimal teacher-student interaction versus non-traditional or alternative instruction, characterized by significant teacher-student interactions. Within this context, traditional instruction typifies teaching that explicitly conveys facts and principles to students with the intention that students will receive this expert knowledge and be evaluated on what they retain according to some preset standards. On the other hand, alternative instruction is synonymous with teaching scientific process skills such as reasoning about problems to find viable solutions, in which case students are assessed based on their individual improvement not just on the subject matter they know. Despite calls in the natural sciences to adopt teaching practices consistent with non-traditional methods, faculty face challenges in the form of situational constraints that favor traditional instruction over alternative instruction in many departments (Henderson & Dancy, 2007). Looking broadly at teaching practices in higher education, Collins and Pratt (2010) espouse a general model of teaching that comprise five elements (teacher, learner, content, context, and ideals) and three relationships (teacher-learner, teacher-content, and learner-content) that they used to develop an instrument to encourage self-reflection. Faculty register their level of agreement or disagreement on a 5-point Likert scale to items such as "Teaching should build on what people already know," and "I encourage people to challenge each other's thinking" on the personal beliefs and intention portions of the instrument, respectively (Collins & Pratt, 2010, p. 363). According to Collins and Pratt (2010), the main purpose of the instrument is to provide a mechanism for faculty to think about their teaching. Collins and Pratt (2010) insist the instrument is not an apparatus that casts judgment about what is good quality teaching or is it a device that specifies criteria for what is effective teaching.

Convinced there was no instrument that efficiently captures the broad spectrum of teaching practices in STEM at postsecondary institutions, Wieman and Gilbert (2014) developed the Teaching Practice Inventory (*TPI*) to determine the extent to which research-based teaching practices are being used in science courses. Wieman and Gilbert (2014) identified eight categories of teaching they reason provide a snapshot of the learning context in typical science courses. They are (1) course information provided, (2) supporting materials, (3) in-class features and activities, (4) assignments, (5) feedback and testing, (6) other teaching practices not previously mentioned, (7) training and guidance for teaching assistants, and (8) collaboration or sharing in teaching. Opting not to determine internal consistency or assess construct validity using standard psychometric procedures, Wieman and Gilbert (2014) concentrated instead on establishing validity based on the consistency and accuracy with which teachers interpreted the items. Wieman and Gilbert (2014) argue the *TPI* is not designed to measure an underlying ability to teach, but rather, to characterize the range of teaching practices in a course. Items about course information and supporting materials provided ask teachers to check all the boxes that apply. Similarly, items about activities and features of tasks performed in class ask teachers to report the number or fraction of times they were done or to check the boxes that describe what was done. Wieman and Gilbert (2014) contend it is futile to measure reliability and construct validity in the traditional sense, since the items do not necessarily correlate within a given category or were they intended to load onto pre-determined latent factors describing aspects of teaching practice. The *TPI* was constructed to provide a dependable and realistic depiction of teaching practices students encounter, and by extension, the learning environment they experience in a science course. This makes the *TPI* particularly fitting to measure the teaching and learning context for this study.

Selection of Psychometric Instruments

The purpose of this study is to operationalize a model of conceptual change for the topic of evolution by natural selection to understand the impact, if any, that constructs formal reasoning ability, academic motivation, and teaching practice have on changing student misconceptions. It is imperative that reliable and valid instruments, appropriate for the scope and goals of this study, are used to measure each construct. The administration of the Conceptual Assessment of Natural Selection test (*CANS*) as a pre- and post-test provides a practical and psychometrically rigorous means for assessing conceptual change in students' understanding of evolution by natural selection. The use of the Montana State University Formal Reasoning Test (*MSU-FORT*) is an efficient and trustworthy way to quantitatively measure students' reasoning about scientific processes and skills necessary to carry out investigations. Self-report questionnaires are widely used to evaluate students' motivation. The Science Motivation Questionnaire II (*SMQ-II*) is a convenient and credible method used to determine students' academic motivation to learn biology. Finally, the Carl Wieman Science Education Initiative Teaching Practice Inventory instrument (*CWSEI_TPI*) evaluates the extent to which teachers report that they incorporate research-based instructional practices in their teaching. In this study, the *CWSEI_TPI* is used to get a sense of the teaching and learning context from the teacher's perspective and to score self-reported teaching practices against normative standards.

Chapter Summary

There are very few empirical studies in the natural sciences that investigate the impact of students' cognitive ability and personal attributes on how likely students are to undergo the process of conceptual change and correct firmly held misconceptions. This is due partly to the lack of congruity in the domains of cognitive psychology and science education, the foremost fields of study that engage in research on how conceptual change occurs and what facilitates conceptual change. The classical conceptual change model originating out of the domain of science education has undergone several modifications to address criticisms of a purely rational perspective to explain and predict the process of conceptual change. Subsequent modifications, for example the extended conceptual change model, incorporates extrarational components to account for differences in students' willingness to change their misconceptions, despite having similar background knowledge. Later models build on prior models. For instance, the cognitive reconstruction of knowledge model adds to the extended model of conceptual change and explores how the persuasiveness of the new information influences the strength and duration of the conceptual change. Operationalizing a model of conceptual change dictates finding reliable and valid instruments that credibly and consistently measure each construct. Instruments that measure constructs of conceptual knowledge of evolution by natural selection, formal reasoning ability, and academic motivation, as well as teachers' perspective of their teaching practice which have sound psychometric properties are available for classroom use. An instrument is selected for each of these constructs that is appropriate for the objectives and scope of this study.

CHAPTER THREE

METHODOLOGY

Introduction

The main purpose of this study is to operationalize a model of conceptual change for the theory of evolution taught in introductory biology courses at public four-year colleges and universities in the United States. More precisely, the objective is to operationalize a model to determine the relationship, if any, among the constructs of change in students' conceptual understanding of evolution by natural selection (conceptual change), formal reasoning ability (cognitive), and academic motivation (attributional), while accounting for the teaching practice students experience (contextual) and controlling for student demographic variables of biological sex, declaration as a biology major, and representation of racial group in science. To bolster the robustness of such an operational model of conceptual change, it is crucial to carry out the study in several educational settings; preferably, a reasonable number of postsecondary institutions across the nation. It is very difficult to conduct a scientific investigation on conceptual change that targets multiple constructs (here, cognitive, attributional, and contextual) at one postsecondary institution, let alone many over a wide geographical area. To do so requires the collective expertise of researchers across academic disciplines and the cooperation of various individuals in different departments at multiple postsecondary institutions, all of which is challenging to undertake and even harder to achieve (Murphy & Alexander, 2008). This study is part of an NSF funded (#1432577) collaboration between ecology and education faculty at Montana State University (MSU). Its conception, execution, and ultimate completion necessitated the strategic and persistent efforts of many persons in two academic disciplines, several departments, and seven postsecondary institutions over a period of 4 to 5 years.

Research Questions

This study addresses the following research questions for undergraduate introductory biology students at six institutions for the topic of evolution by natural selection:

1. To what extent do students' formal reasoning ability and academic motivation predict levels of conceptual change?
2. Which variable, or combination of variables, that is, formal reasoning ability, academic motivation, teaching practice, teacher experience, and student demographics, are most likely to predict levels of conceptual change?

In addition to the dependent variable of conceptual change and the independent variables of students' formal reasoning ability and academic motivation, a set of control variables of students' demographic information as well as the teaching practice and teacher experience students encounter are included in the hierarchical linear modeling and hierarchical regression analyses.

Role of Researcher

I was hired as a graduate research assistant to work on this NSF (#1432577) project. My primary role was to coordinate the recruitment of biology faculty and handle the logistics of carrying out a scientific investigation at multiple sites. I was responsible for compiling a list of postsecondary institutions that comprise the population in the study. I was also charged with selecting a random sample of institutions from the population, contacting biology faculty assigned to teach the introductory biology course intended for majors at the sampled institutions, and encouraging prospective introductory biology instructors to participate in the study. Once an introductory biology instructor agreed to participate, I reached out to the person(s) in charge of overseeing research on human subjects for approval to conduct the study at the participating institution. I worked on behalf of the principal investigator of the NSF project at MSU and the participating instructor to liaise with the Institutional Review Board (IRB) at the participating site to submit all required documentation and procure all necessary permissions from the participating department (and college where applicable). When all the mandatory paperwork was submitted and approved, I arranged with the participating introductory biology instructor how, when, and where students would complete each of the instruments required for the study. Also, I discussed and confirmed with the participating introductory biology instructor what he or she needed to do and what obligations must be met for the successful completion of the study. This included coordinating the dissemination of all instruments and the collection of all data for both students and the instructor at each participating site. I subsequently analyzed the data and sent an aggregate report of students' performance on the pre- *CANS* and post-*CANS* to each participating introductory biology instructor. I also initiated the process for dispensing the stipend to each introductory biology instructor who satisfactorily met the requirements for participating in the study.

Rationale for Research Method

This quantitative study has a pre-test/post-test design to operationalize a model of conceptual change to explain change in conceptual understanding of evolution by natural selection and to describe the relationship, if any, among formal reasoning ability and academic motivation, accounting for teaching practice and teacher experience for introductory biology students at six postsecondary institutions. For this investigation, the pre-test was administered to students before instruction and the post-test after instruction on the unit of evolution. The pre-test and the post-test scores are paired for each student. The change or difference in scores is attributed to change in understanding of evolution by natural selection. The pre-test/post-test design to study change in knowledge has a long tradition in quantitative research (Bonate, 2000; Cronbach & Furby, 1970; Dimitrov & Rumrill, 2003; Williams & Zimmerman, 1996). But from as far back as the 1950s to early 1970s, there has been a debate over the reliability of change or difference scores for paired data on the same subject in quantitative analyses. This disagreement has sparked sharp reactions from psychometricians with a notable and widely regarded assertion from Cronbach and Furby (1970) that difference scores are not as reliable as the pre-test and post-test scores themselves, and therefore the use of difference scores should not be encouraged in regression analyses. Yet, over the years, researchers have questioned the underlying premise for the unreliability of difference scores (Williams & Zimmerman, 1996), contending this assertion is only true if the pre-test and post-test scores have equal variance and equal reliability (Bonate, 2000; Dimitrov & Rumrill, 2003). Regardless, the prevailing practice is to use the post-test score as the dependent variable and pre-test score as a covariate; that is, an independent variable in the regression analyses (Bonate, 2000; Dimitrov & Rumrill, 2003). This approach minimizes regression to the mean, and it side steps the dispute that often arise with regards to difference scores (Bonate, 2000). Therefore, in this study, to examine change in conceptual understanding for evolution by natural selection in the regression analyses, the post-test score is the dependent variable, and the pre-test score, formal reasoning ability score, and academic motivation scores are independent variables, plus student demographics, teaching practice, and teacher experience are the control variables.

This study involves introductory biology students at six postsecondary institutions (*SCHOOL_A*, *SCHOOL_B*, …, *SCHOOL_F*) in the United States. Accordingly, there are two levels of data: level one for the observational unit of each student, and level two for the teaching practice and teacher experience of the introductory biology instructor contextual at each school (see Figure 3.1 below on page 82). To account for the nested structure of the data, hierarchical linear models with a random effect for the school attended is used to address the possible dependence or intercorrelation of student scores within each postsecondary institution (Galecki & Burzykowski, 2013; Pinheiro, Bates, DebRoy, Sarkar, & R Core Team, 2018; Zuur, Ieno, Wlaker, Saveliev, & Smith, 2009). The regression analyses are carried out on the dependent (response) variable of each student's post-test score and the fixed effects (explanatory variables) of each student's pre-test score, formal reasoning ability score, and academic motivation score. This study also controls for the fixed effects of students' biological sex, declaration as a biology major, and representation of racial group in science, plus the teaching practice and teacher experience of the instructor at the postsecondary institution A random intercept is included to account for students enrolled in the same school. In addition to hierarchical linear modeling analyses, a hierarchical linear regression analysis is conducted to determine the most parsimonious model of conceptual change (Akaike, 1974; Bartoń, 2018; Burnham & Anderson, 2004; Burnham, Anderson, & Huyvaert, 2011). That is, the data is examined to find which combination of variables (pre-test, formal reasoning ability, academic motivation, biological sex, declaration as a biology major, representation of racial group, teaching practice, and/or teacher experience) best explains the post-test results for the conceptual understanding of evolution by natural selection in this study on introductory biology students at the six postsecondary institutions across the nation.

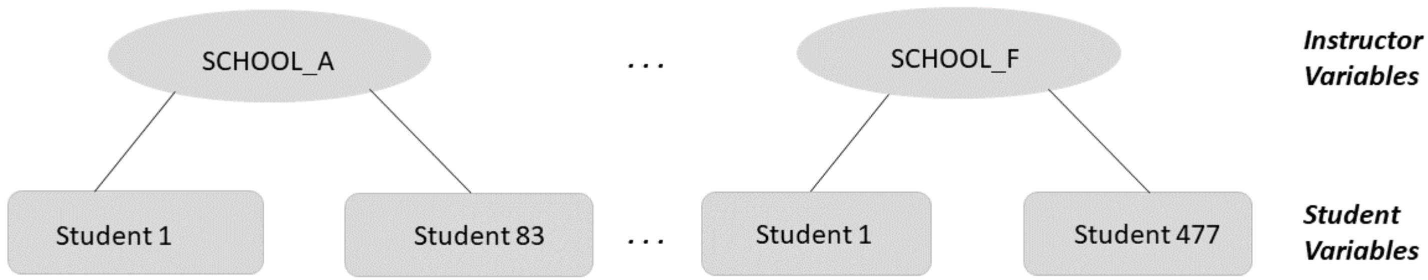


Figure 3.1. Graphical representation of the structure of the data

Data Collection Procedures

Every effort was made to obtain a representative sample of introductory biology students at public four-year colleges and universities across the United States. From a list of 258 public four-year colleges and universities in the nation that offer a baccalaureate degree in biology and have 10,000 or more students, 30 institutions were randomly selected to participate in the study. Instructors of sections of the introductory biology course intended for biology majors whose content covers the unit of evolution were contacted and invited to participate. Of the 30 institutions randomly selected, six decided to participate. IRB approval was required for the principal investigator's institution and each participating site. Also, on average, a total of 80-90 minutes of students' time was needed over the course of the semester to carry out the study. An introductory biology instructor at each participating institution agreed to undertake the IRB application process and set aside the time needed, either in class, during lab sessions, or in the case on one institution, as an out-of-class online assignment for full participation in the study. If there were multiple sections with different instructors, only one instructor at each institution agreed to the conditions. This worked to our advantage since budgetary constraints led to a research design that asked for one instructor at each institution. Instructors received a stipend for their participation to offset the cost and time spent performing the tasks necessary to conduct the study.

Instructors were asked to administer all instruments to their students over the duration of the unit of evolution, spanning two to three weeks, or the entire semester if preferred. Typically, the unit of evolution is taught at the beginning of the course and reinforced and applied throughout the semester. Some instructors chose to complete the study during the unit of evolution, especially if the course is taught by more than one instructor. Other instructors who teach the whole course opted to conduct the study over the entire semester. To carry out the study, instructors were asked to administer a pre-test assessing students' knowledge of evolution by natural selection before instruction begun on the unit of evolution, a post-test after instruction was done, and a third combined instrument testing students' formal reasoning ability and documenting students' academic motivation during the unit of evolution, or just after the unit was finished. Students were also asked to complete select demographic questions at the end of the pre-test. Scheduling restrictions necessitated some flexibility for the administration of the third combined instrument. Students at five institutions completed the instruments using pencil and SCANTRON® Optical Mark Recognition (OMR) scan forms, and students at one institution completed the instrument using the MSU licensed Qualtrics LLC online survey platform. The completed SCANTRON® OMR scan forms were mailed and the Qualtrics LLC files downloaded to the principal investigator's institution for scoring and analyses. Instructors completed a teaching practice inventory before instruction began on the unit of evolution. They were also asked to provide information about how many semesters they have taught the introductory biology course as a member of faculty. For their convenience, instructors responded to the questions using the MSU Qualtrics LLC online survey platform. Once all the instructors completed instrument, the Qualtrics LLC file was downloaded to the principal investigator's institution for scoring and analyses.

Description of Sample

A summary of the description for the sampled population of introductory biology students at the six participating postsecondary institutions (R Core Team, 2016) is given in Table 3.1 (page 88). Class sizes ranged from 125 to 1061 students, and the number of students who fully participated in the study and provided a complete data set was from a low of 59 with a corresponding response rate of 47% at *SCHOOL_C* to a high of 477 and a response rate of 45% at *SCHOOL_F*. The highest participation was for *SCHOOL_E* with 116 of 131 students responding at a rate of 89%. Students at all six schools were chiefly between the ages of 18-21 years. The percentage of students in this age group ranged from 82%-99% at the participating schools. The proportion of males and females, respectively, were close to equal for two institutions, *SCHOOL_B* (48%, 52%) and *SCHOOL_D* (48%, 52%); but there were less males than females for the remaining four institutions, *SCHOOL_A* (35%, 65%), *SCHOOL_C* (36%, 64%), *SCHOOL_E* (29%, 71%), and *SCHOOL_E* (25%, 75%). Although the study targeted introductory biology courses for majors, there were instances when the course comprised of less majors than nonmajors because the study took place in a semester when most of the majors in that cohort had already completed the course in a previous semester. Another reason too, there just happened to be less biology majors in that section of the course for that instructor. In any case, the proportion of biology majors ranged from as low as 7% for *SCHOOL_B* to as high as 85% for *SCHOOL_C*. Every school in the study had a greater proportion of students who identified with a racial group that was in the majority for representation in science. Only *SCHOOL_A* had a proportion under 70% majority representation in science. In all, complete data for the study variables were obtained for 1140 students of the 2064 students enrolled in the introductory biology courses at the six postsecondary institutions, corresponding to an overall response rate of 55%. Consistent with data from the individual schools, for the combined sample of introductory biology students from the six schools, most students (92%) were between 18-21 years old, the majority (65%) were females, a minority (35%) were biology majors, and a much greater proportion (82%) identified with a racial group that was in the majority of persons represented in science.

Two pieces of evidence are presented to ascertain whether the study sample is representative of the target population of introductory biology students in the United States at public four-year colleges and universities that offer a bachelor's degree in biology with 10,000 or more students. First, a profile of the six institutions that participated in the study is given in Table 3.2 (page 89). Information on the faculty-to-student ratio, Carnegie classification as a research institution, in addition to students' retention rate, graduation rate, and estimated total on campus expenses are offered to help determine how typical these institutions are when compared to other institutions across the nation (Indiana University Center for Postsecondary Research [IUCPR], 2017; NCES, 2018). Second, a comparison is made between the course grades for introductory biology students who fully participated in the study and those who partially or did not participate to evaluate if there is a difference between the two groups. There was only one introductory biology instructor who taught the unit of evolution for three of the six institutions. For the other three institutions, only one of the introductory biology instructors assigned to teach the unit of evolution participated in the study. Students' course grades were available for five of the six institutions. The Welch *t*-test and Mann-Whitney Wilcoxon rank sum and signed rank test were performed to evaluate if there is a difference in the mean and median course grades, respectively, between the students who fully participated in the study and those who partially or did not participate for five institutions (R Core Team, 2016). The results are summarized in Table 3.3 (page 90). For all schools except *SCHOOL_E*, there is strong evidence there is a difference between the mean course grades of the students who fully participated and those who partially or did not participate in the study. Students who fully participated tend to have a higher mean course grade than students who partially or did not participate in the study. For all schools, there is moderate to strong evidence there is a difference between the distribution of course grades, and the median scores, for the students who fully participated and those who partially or did not participate in the study. The median course grade of students who fully participated is greater than students who partially or did not participate in the study for five of six schools. Therefore, it is reasonable to conclude that, in general, students with higher course grades were more likely to complete all the instruments and provide a complete data set in the study.

Table 3.1. Frequency summary of descriptive data for participating schools and all schools combined

VARIABLE	*SCHOOL_A*	*SCHOOL_B*	*SCHOOL_C*	*SCHOOL_D*	*SCHOOL_E*	*SCHOOL_F*	TOTAL
Class Size	138	316	125	292	131	1061	2064
Sample Size	83	205	59	200	116	477	1140
Response Rate	0.60	0.65	0.47	0.68	0.89	0.45	0.55
Age							
18-21	0.86	0.84	0.92	0.82	0.99	0.99	0.92
22-25	0.07	0.09	0.07	0.12	0.01	0.01	0.05
>25	0.07	0.07	0.02	0.05	0.00	0.00	0.03
Sex							
Male	0.35	0.48	0.36	0.48	0.29	0.25	0.35
Female	0.65	0.52	0.64	0.52	0.71	0.75	0.65
Biology Major							
Yes	0.58	0.07	0.85	0.56	0.58	0.24	0.36
No	0.42	0.93	0.15	0.44	0.42	0.76	0.64
Representation in Science							
Majority	0.57	0.94	0.75	0.72	0.86	0.86	0.82
Minority	0.43	0.06	0.25	0.28	0.14	0.14	0.18

Note: Proportions reported for levels of categorical variables Age, Sex, Biology Major, and Representation in Science
Majority – White and Asian
Minority - Black, American Indian or Alaska Native, and other groups

Table 3.2. A profile of institutions that participated in the study

	SCHOOL_A	*SCHOOL_B*	*SCHOOL_C*	*SCHOOL_D*	*SCHOOL_E*	*SCHOOL_F*
Faculty: Student Ratio	20:1	15:1	18:1	22:1	15:1	19:1
Carnegie Classification[a]	Doctorial Universities: Moderate Research Activity (R3)	Doctorial Universities: Higher Research Activity (R2)	Doctorial Universities: Higher Research Activity (R2)	Doctorial Universities: Highest Research Activity (R1)	Doctorial Universities: Highest Research Activity (R1)	Doctorial Universities: Highest Research Activity (R1)
Retention Rates for first-time students[b]						
Full-time	69%	78%	71%	85%	91%	96%
Part-time	43%	35%	8%	61%	100%	81%
Graduation Rates for full-time students (completion)[c]						
4 years	16%	26%	28%	25%	73%	68%
6 years	37%	58%	48%	54%	83%	88%
Estimated total expenses for full-time beginning undergraduate students in USD (on campus)[d]						
In-state	21,294	19,777	24,067	24,946	27,792	21,131
Out-of-state	32,080	31,387	35,653	37,306	47,782	43,409

Note: USD – United States dollars

[a] Carnegie Classification – Basic

[b] Retention Rates for first-time students – pursuing bachelor's degrees who return the following fall

[c] Graduation Rates for full-time students – pursuing bachelor's degrees who began Fall 2011

[d] Estimated total expenses for full-time beginning undergraduate students in USD (on campus) – for academic year 2017-2018

Table 3.3. A comparison of the mean and median course grade scores for students who fully participated and those who partially or did not participate in the study

	SCHOOL_A	*SCHOOL_B*	*SCHOOL_C*	*SCHOOL_D*	*SCHOOL_E*	*SCHOOL_F*
Mean % score based on participation in study						
Full	75.8	-	79.3	85.9	84.2	80.6
Partial/None	52.3	-	71.2	74.1	79.0	76.4
Full – Partial/None	23.5	-	7.9	11.8	5.2	4.2
t	6.02	-	3.85	7.78	1.81	6.11
df	80.12	-	119.25	124.96	11.07	1011.7
p-value	<0.0001	-	<0.0001	<0.0001	0.098	<0.0001
95% *CI*	(15.7, 31.3)	-	(3.8, 12.0)	(8.8, 14.8)	(-1.1, 11.5)	(2.8, 5.5)
Median % score based on participation in study						
Full	77.4	-	79.0	85.0	84.5	80.8
Partial/None	52.8	-	71.8	75.0	77.7	78.2
Wilcoxon difference	22.6	-	7.6	10.0	6.2	3.2
W	3934.5	-	2669	14476	918	166250
p-value	<0.0001	-	0.0004	<0.0001	0.0441	<0.0001
95% *CI*	(12.0, 34.1)	-	(3.5, 12.0)	(10.0, 15.0)	(0.3, 11.4)	(2.0, 4.4)

Course grades were not available for *SCHOOL_B*
t –Welch two sample *t*-test (assumes variance of full and partial/none participation are not equal)
Wilcoxon difference – difference in scores between the two groups based on Wilcoxon rank sum ranks
W – Wilcoxon sum of the ranks of the scores
df – degrees of freedom
p-value – two-tailed
95% *CI* – 95% confidence interval of the difference between full and partial/none participation (full – partial/none)

Operational Definition of Study Variables

In this study, conceptual change is operationalized as students' conceptual understanding of evolution by natural selection assessed as the number of accurate responses on the Conceptual Assessment of Natural Selection test (*CANS*, Kalinowski et al., 2016 with the revised 20-item version used in this study exhibited in Appendix A), before instruction (*PRE*) and after instruction (*POST*). Students' formal reasoning ability (*FORMAL*) is measured as the total correct answers on the Montana State University Formal Reasoning Test (*MSU-FORT*, Kalinowski et al., 2018 with the modified 22-item version used in this study included in Appendix B). Students' self-reported motivation to learn biology as it relates to their self-efficacy or beliefs about their capabilities (*SELF_EFF*), self-determination or control over events related to learning (*SELF_DET*), interest in learning for personal satisfaction and subject mastery (*INTRINSIC*), grade they want to achieve (*GRADE*), and intended career (*CAREER*) are quantified using a Likert scale with 1 being not at all favorable to 5 being extremely favorable on the Science Motivation Questionnaire II (*SMQ-II*, Glynn et al., 2011 given in Appendix C). Instructors completed the Carl Wieman Science Education Initiative Teaching Practice Inventory instrument (*CWSEI_TPI*, Wieman & Gilbert, 2014 shown in Appendix D). Their *CWSEI_TPI* responses were scored using a set of rules that reflected the instructor's self- reported level of research-based teaching practices against normative standards (Wieman & Gilbert, 2014). The proportion of the total possible points was reported for each instructor. Instructors also reported the number of times they taught the introductory biology course as a faculty member (*Semesters_Taught*). Students reported their biological sex (*f.SEX*), whether they were a biology major (*f.MAJ_BIO*), and the race with which they identify. White students and students of Asian ethnicity were combined as a group with majority representation in science, and all other racial groups were combined as having minority representation in science (*f.RACE*). To summarize, this study operationalizes a model of conceptual change with dependent quantitative variable, *POST*, and independent quantitative variables *PRE*, *FORMAL*, *SELF_EFF*, *SELF_DET*, *INTRINSIC*, *GRADE*, and *CAREER* at the student level; control categorical variables *f.SEX*, *f.MAJ_BIO*, and *f.RACE* at the student level; in addition to control quantitative variables *CWSEI_TPI* and *Semesters_Taught* at the school level. Table 3.4 (pages 93-96) lists each variable, the measurement scale for each variable, the value or code used for each variable, and a sample item for each variable from the respective instrument (Mills & Gay, 2016).

Table 3.4. Variable, variable measurement scale, variable value or code, and variable sample item from instrument

Variable	Scale	Instrument (total items)	Value/Code	Sample item (numbered label on instrument)
Response				
POST	Ratio	*CANS* (20)	Number of correct responses	What traits do saguaro cacti inherit from their parents? a. Traits that helped their parents survive and reproduce. b. Traits that were changed by the environment during their parents' lifetime. c. Traits that were determined by genes. d. Traits determined by genes plus one or more other traits listed above. (#5)
Explanatory				
PRE	Ratio	*CANS* (20)	Number of correct responses	What traits do saguaro cacti inherit from their parents? a. Traits that helped their parents survive and reproduce. b. Traits that were changed by the environment during their parents' lifetime. c. Traits that were determined by genes. d. Traits determined by genes plus one or more other traits listed above. (#5)
FORMAL	Ratio	*MSU-FORT* (22)	Number of correct responses	Brendan wants to know how trees move water from their roots up to their leaves. One potential explanation is that cells in the roots have molecular pumps that push water upwards. Another explanation is that leaves have molecular pumps that suck water upwards. Brendan cuts the roots off several small trees and places the trees in buckets of water containing red food

Variable	Scale	Instrument (total items)	Value/Code	Sample item (numbered label on instrument)
				coloring. An hour later, he observes that the water containing the red food coloring has risen to the top of these trees. What can Brendan conclude? a. Molecular pumps in roots are pushing water to the top of his trees. b. Molecular pumps in leaves are sucking water to the top of his trees. c. Molecular pumps in roots are not responsible for moving water up his trees. d. He cannot conclude much because the trees in his experiment did not have roots. e. He cannot conclude much because his experiment did not have a control. (#15)
SELF_EFF	Interval	*SMQ-II* (5)	Integer: 1 (*never*) to 5 (*always*)	I am confident I will do well on biology tests. (#9)
SELF_DET	Interval	*SMQ-II* (5)	Integer: 1 (*never*) to 5 (*always*)	I prepare well for biology tests and labs. (#16)
INTRINSIC	Interval	*SMQ-II* (5)	Integer: 1 (*never*) to 5 (*always*)	Learning biology makes my life more meaningful. (#12)
GRADE	Interval	*SMQ-II* (5)	Integer: 1 (*never*) to 5 (*always*)	Getting a good biology grade is important to me. (#4)
CAREER	Interval	*SMQ-II* (5)	Integer: 1 (*never*) to 5 (*always*)	My career will involve biology. (#23)
f.SEX	Nominal	*Demographics*	Male Female	What is your sex? a. Female b. Male

Variable	Scale	Instrument (total items)	Value/Code	Sample item (numbered label on instrument)
f.MAJ_BIO	Nominal	*Demographics*	Major (majoring and minoring) Nonmajor (not pursuing major or minor)	Are you pursuing a degree in Biology? a. I am majoring in Biology b. I am minoring in Biology c. I am not pursing a major or a minor in Biology
f.RACE	Nominal	*Demographics*	Majority (White or Asian) Minority (All other race categories)	What category best describes your race? a. American Indian or Alaska Native b. Asian c. Black or African American d. White e. Another category
Semesters_Taught	Ratio	*Demographics*	Number of semesters taught course	How many times have you taught this course as a faculty member?
CWSEI_TPI	Ratio	*CWSEI_TPI* (8 sections; Parts I-VIII)	Proportion of maximum score for best practice	Which of the following assignments do you use? Check all that apply in your course. ☐ Problem sets/homework assigned or suggested but did not contribute to course grade ☐ Problem sets/homework assigned and contributed to course grade at intervals of 2 weeks or less ☐ Paper or project (an assignment taking longer than two weeks and involving some degree of student control in choice of topic or design) [3] ☐ Encouragement and facilitation for students to work collaboratively on their assignments ☐ Explicit group assignments ☐ Other (please specify below).

Variable	Scale	Instrument (total items)	Value/Code	Sample item (numbered label on instrument)
				If you selected Other above, please specify here. ____________________ (Part IV)

Description of Data

Table 3.5 (page 100) reports the mean (*M*) and standard deviation (*SD*) for students' post-instruction, pre-instruction, formal reasoning ability, and academic motivation scores, plus instructors' teaching practice score and the number of semesters they taught the introductory biology course (R Core Team, 2016). On average, in all six schools, students' understanding of evolution by natural selection increased after instruction on the unit of evolution. Before instruction, students got most of the 20 questions on the *CANS* incorrect with scores from *M*=4.4; *SD*=4.0 for *SCHOOL_A* to *M*=7.5; *SD*=4.6 for *SCHOOL_C*. After instruction, students got less than half to about half of the 20 questions correct on the *CANS* with scores from *M*=6.2; *SD*=4.6 for *SCHOOL_A* to *M*=10.6; *SD*=5.3 for *SCHOOL_E*. This trend of increased understanding after instruction is reflected in the overall sample of introductory biology students with before instruction scores *M*=6.4; *SD*=4.5 and after instruction scores *M*=8.8; *SD*=4.9. Students' formal reasoning ability scores ranged from *M*=10.3; *SD*=3.5 for *SCHOOL_A* to *M*=13.3; *SD*=3.6 for *SCHOOL_F*, which is about half to just above half of the 22 questions correct. For the full sample of introductory biology students, the formal reasoning score was *M*=13.3; *SD*=3.6. Across all six schools, the sample of introductory biology students reported they had a favorable perception of their self-efficacy to learn biology (*M*=3.8; *SD*=0.7), self-determination to exert control over their learning circumstance (*M*=3.8; *SD*=0.7), and intrinsic or personal interest in learning biology (*M*=3.7; *SD*=0.8) on a 5-point Likert scale. Students also reported they had a very favorable perception of their extrinsic motivation to learn biology as it relates to their course grade (*M*=4.5; *SD*=0.6) and future career (*M*=4.0; *SD*=1.0). Instructors self-reported on their use of research-based teaching practices in the introductory biology course, which was converted to a proportion scale corresponding to 0 (*none*) to 1 (*exemplary*) levels of best practice based on normative standards. Individual scores ranged from 0.33 to 0.70. As a group, the six instructors' scores were *M*=0.46; *SD*=0.13. There was great variation in teaching experience among instructors with times teaching the introductory biology course ranging from 8 to 38 and average *M*=8.4; *SD*=9.3.

Stacked histograms for the distribution of the data for all schools combined, and partitioned by school (Arnold, 2017; Pruim, Kaplan, & Horton, 2017; Wickham, 2007, 2009; Wickham & Bryan, 2017), are displayed in Figure 3.2 (page 101) and Figure 3.3 (page 102). The pre-instruction, post-instruction, learning gain or difference scores (*POST – PRE*), and formal reasoning scores are shown in Figure 3.1, and the motivation scales are displayed in Figure 3.2. For the complete sample of introductory biology students, the distribution of the pre-instruction scores is right-skewed with a noticeable shift towards more students getting more questions correct after instruction for the post-instruction scores. The learning gain is almost normally distributed. Also, the distribution for the formal reasoning scores is approximately normal. In comparison, the motivation ratings on all five scales are left-skewed, and this is particularly evident for the career and grade motivation scales. For a detailed look at each school, the distributions of the pre-instruction, post-instruction, formal reasoning, and motivation data are presented in Appendix E. The first slide compares superimposed density plots of the pre-instruction and post-instruction scores for each school. The second slide displays histograms of the formal reasoning scores for each school. The third slide shows a grid for each school with side-by-side boxplots of each component of motivation (Arnold, 2017; Pruim, Kaplan, & Horton, 2017; Wickham, 2007, 2009; Wickham & Bryan, 2017).

Table 3.5. Means and standard deviations of quantitative variables for participating schools and all schools combined

VARIABLE	*SCHOOL_A*	*SCHOOL_B*	*SCHOOL_C*	*SCHOOL_D*	*SCHOOL_E*	*SCHOOL_F*	TOTAL
STUDENT							
Post-Instruction *(correct of 20)*	6.2(4.6)	8.2(4.6)	9.9(4.6)	8.7(4.8)	10.6(5.3)	9.1(4.7)	8.8(4.9)
Pre-Instruction *(correct of 20)*	4.4(4.0)	5.2(4.0)	7.5(4.6)	6.6(4.6)	7.0(5.1)	6.9(4.4)	6.4(4.5)
Formal Reasoning *(correct of 22)*	10.3(3.5)	11.4(3.3)	12.6(3.6)	12.7(3.7)	12.4(3.0)	13.3(3.6)	12.5(3.6)
Motivation *(5-Point Scale)*							
Self-Efficacy	4.1(0.7)	3.9(0.6)	4.1(0.6)	3.8(0.8)	3.8(0.7)	3.7(0.8)	3.8(0.7)
Self-Determination	3.7(0.8)	3.8(0.6)	4.0(0.6)	3.6(0.8)	3.9(0.7)	3.8(0.7)	3.8(0.7)
Intrinsic	3.7(0.9)	3.6(0.8)	4.1(0.6)	3.8(0.9)	3.8(0.8)	3.6(0.8)	3.7(0.8)
Grade	4.6(0.6)	4.4(0.6)	4.6(0.4)	4.4(0.6)	4.6(0.4)	4.5(0.5)	4.5(0.6)
Career	3.9(1.2)	3.9(1.0)	4.5(0.5)	4.0(1.0)	4.2(0.8)	4.0(0.9)	4.0(1.0)
INSTRUCTOR							
CWSEI TPI† *(proportion)*	0.46	0.46	0.70	0.64	0.60	0.33	0.46(0.13)
Semesters Taught	20	17	38	29	10	8	8.4(9.3)

†Proportion of total score (70)

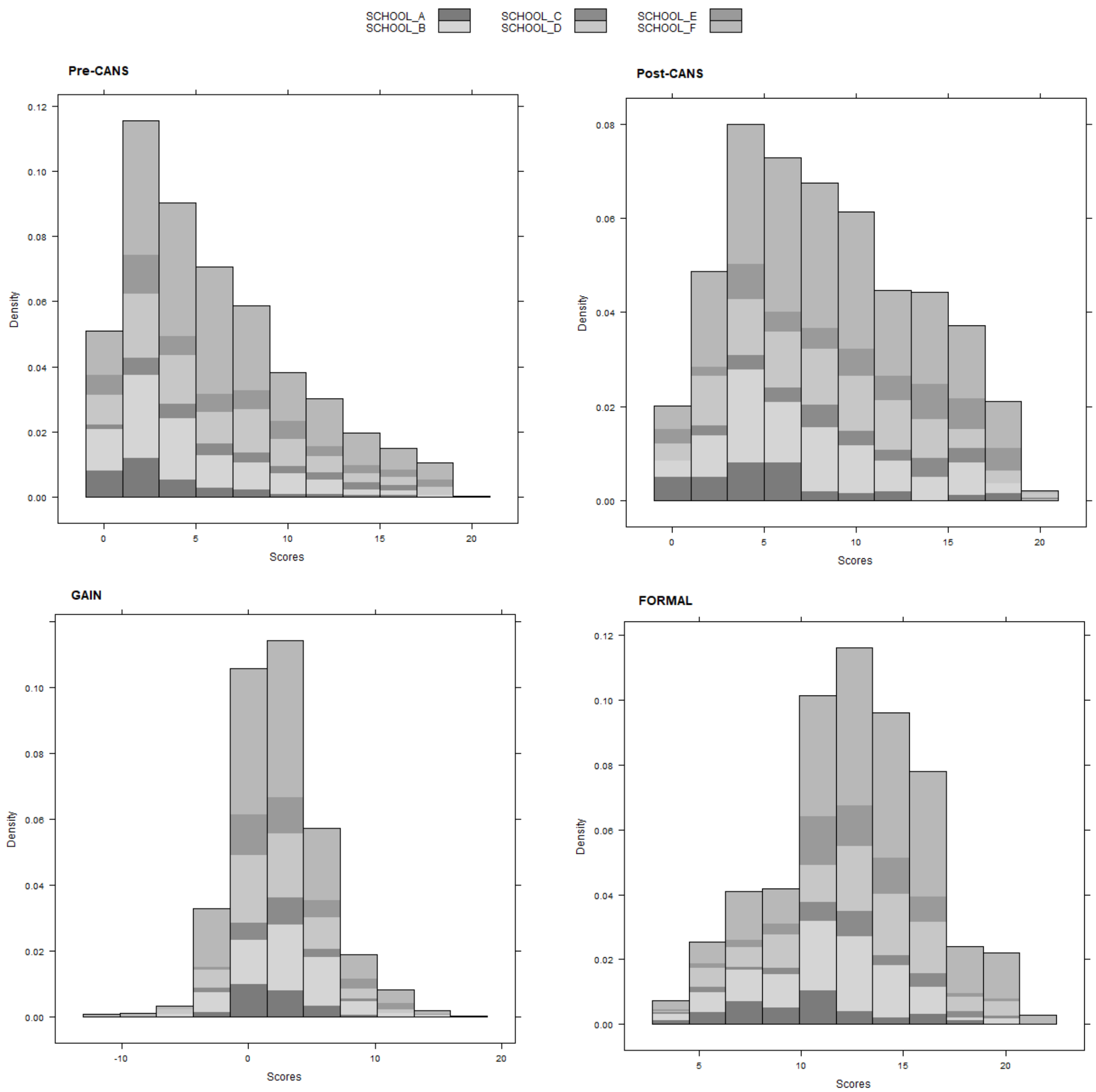


Figure 3.2. Distribution of pre-instruction, post-instruction, learning gain, and formal reasoning scores for all schools combined

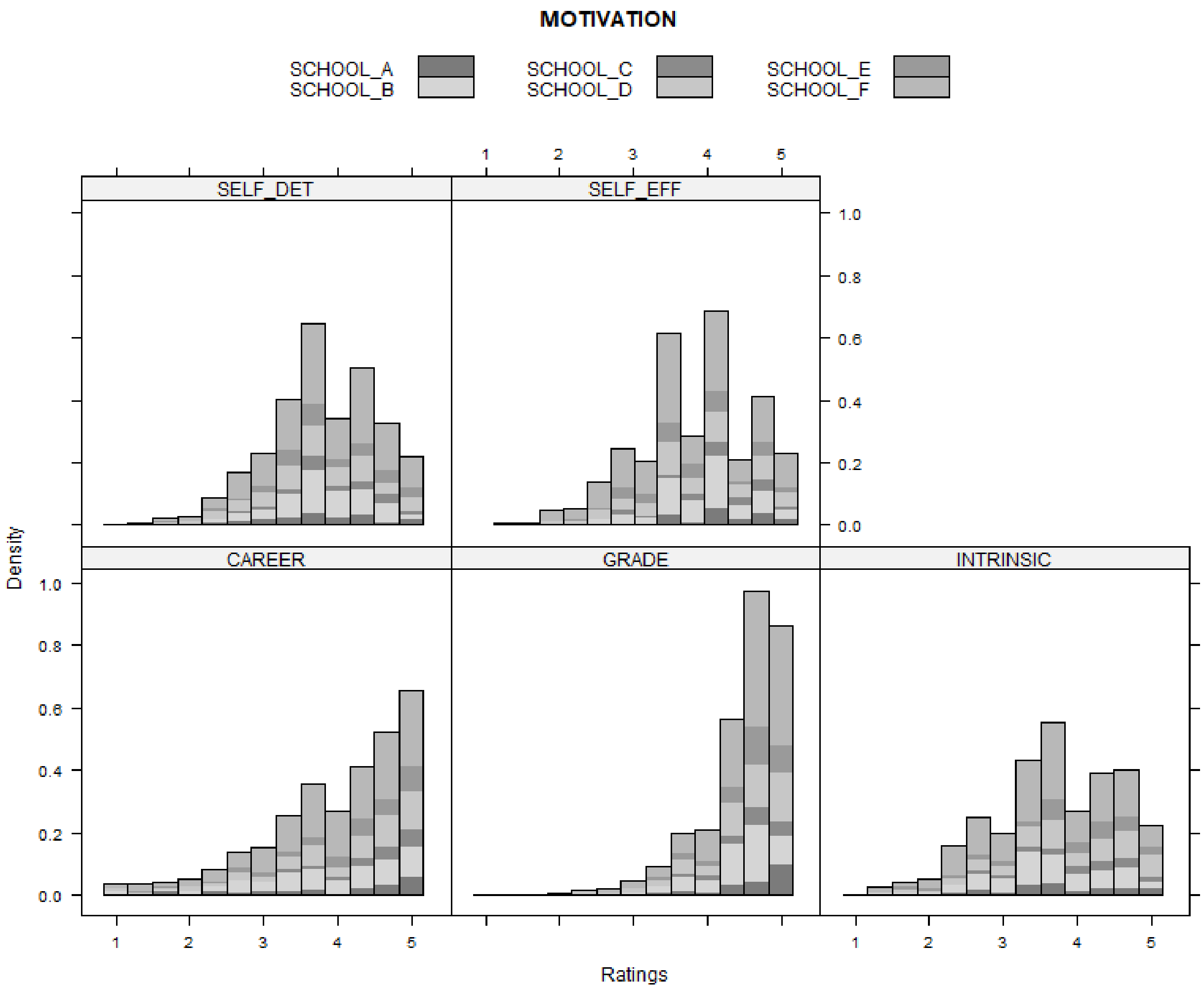


Figure 3.3. Distribution of ratings by component of motivation for all schools combined

Reliability and Validity of Instruments

Cronbach's alpha (α) reliability coefficient is a well-established metric for evaluating the internal consistency of items that attempt to measure the same variable. The estimate of α with the 95% confidence interval (Revelle, 2017) is reported in Table 3.6 (pages 106-107) for each variable measured by the respective instruments in Appendix A (*CANS; Pre*- and *Post*-), Appendix B (*MSU-FORT*), and Appendix C (*SMQ-II*). The *CANS* and *MSU-FORT* are constructed to measure one variable, respectively; that is, all the items on the *CANS* measure students' ability to understand the concept of evolution by natural selection, and in the case of *MSU-FORT*, all items measure students' ability to use formal reasoning. *SMQ-II* has five motivation scales each measuring a distinct component of students' motivation: *SELF_EFF*, *SELF_DET*, *INTRINSIC*, *GRADE*, and *CAREER*. Generally, estimates of α for the complete sample of introductory biology students are consistent with values for the individual schools. For *Pre-CANS* and *Post-CANS*, estimates of α are $0.84 \leq \alpha \leq 0.90$ for the six schools. For the full introductory biology sample, α=0.86 and α=0.87 for *Pre-CANS* and *Post-CANS*, respectively. Estimates of α are lower for *MSU-FORT* with $0.61 \leq \alpha \leq 0.72$ for the six schools, and α=0.71 for the entire sample of introductory biology students. For the motivation variables, estimates of α are similar for all six schools with $0.80 \leq \alpha \leq 0.85$ for *SELF_EFF*, $0.84 \leq \alpha \leq 0.89$ for *SELF_ DET*, $0.84 \leq \alpha \leq 0.89$ for *INTRINSIC*, $0.77 \leq \alpha \leq 0.84$ for *GRADE*, and $0.87 \leq \alpha \leq 0.94$ for *CAREER*. Corresponding estimates of α for the total introductory biology sample are *SELF_EFF* (0.82), *SELF_DET* (0.86), *INTRINSIC* (0.87), *GRADE* (0.80), and *CAREER* (0.92). The reliability coefficient estimates of the study variables for the sample of students in each school and the sample of introductory biology students for all six schools combined are comparable to the internal consistencies published for the respective instruments. A complete data set of all items answered on all instruments administered in the study was available for 1077 introductory biology students from a possible 2064 enrolled in the six schools.

Confirmatory factor analysis (CFA) using structural equation modeling was conducted to assess construct validity for the instruments used in the study (Rosseel, 2012), and the results summarized in Table 3.7 (pages 108-109). The maximum likelihood estimator was used to determine the standardized factor loadings for a latent factor structure for each instrument. For the *Pre-CANS* and *Post-CANS* instruments, all items are constructed to measure a single latent factor of students' conceptual understanding of evolution by natural selection. Similarly, all items on the *MSU-FORT* instrument are designed to measure the latent factor of students' formal reasoning ability. In the case of the *SMQ-II* instrument, 5 items each are intended to measure the latent factors of self-efficacy, self-determination, intrinsic motivation, grade motivation, and career motivation, respectively. Standardized factors loading of ≥0.30 are considered acceptable. For the full sample of introductory biology students, standardized factor loadings were quite reasonable with all items having loadings of ≥0.30 on the expected latent factor for the *Pre-CANS* and *SMQ-II* instruments. Except for 1 item with standardized factor loading of 0.256, loadings of ≥0.30 were obtained for the *Post-CANS*. For the *MSU-FORT* instrument, 8 of the 22 items had factor loadings below 0.30, with an item on logic reasoning having the lowest value of -0.265. Four measures of model fitness are reported for CFA: chi-square (χ^2), root mean square error of approximation (RMSEA), comparative fit index (CFI), and standardized root mean square residual (SRMR), in addition to the ratio chi-square/degrees of freedom (χ^2/df). Desired values are χ^2/df ($\leq$5), CFI ($\approx$0.9), RMSEA ($\geq$0.05, $\leq$0.10), and SRMR ($\approx$0). Based on these measures of model fitness reported in Table 3.7, the overall factor loadings for construct validity is satisfactory and consistent with published results for each instrument. CFA was carried out on the complete data set of all items answered on all instruments for 1077 of 2064 introductory biology students enrolled.

Table 3.6. Cronbach's alpha (α) and 95% confidence interval for the constructs of each instrument by school and all schools combined

INSTRUMENT	*SCHOOL_A*	*SCHOOL_B*	*SCHOOL_C*	*SCHOOL_D*	*SCHOOL_E*	*SCHOOL_F*	TOTAL
Sample Size	81	210	63	200	113	410	1077
Pre-CANS							
Estimate	0.87	0.85	0.86	0.87	0.90	0.84	0.86
Lower	0.84	0.83	0.81	0.85	0.87	0.82	0.85
Upper	0.91	0.88	0.91	0.90	0.92	0.87	0.88
Post-CANS							
Estimate	0.87	0.86	0.84	0.86	0.90	0.85	0.87
Lower	0.83	0.84	0.78	0.84	0.87	0.83	0.86
Upper	0.91	0.89	0.90	0.89	0.93	0.87	0.88
MSU-FORT							
Estimate	0.67	0.66	0.69	0.72	0.61	0.71	0.71
Lower	0.57	0.60	0.58	0.66	0.50	0.67	0.69
Upper	0.77	0.73	0.80	0.77	0.71	0.75	0.74
SMQ-II							
SELF-EFFICACY							
Estimate	0.80	0.81	0.83	0.85	0.80	0.83	0.82
Lower	0.73	0.76	0.77	0.81	0.74	0.80	0.81
Upper	0.87	0.85	0.90	0.88	0.86	0.86	0.84
SELF-DETERMINATION							
Estimate	0.84	0.84	0.85	0.89	0.87	0.85	0.86
Lower	0.78	0.81	0.79	0.86	0.83	0.83	0.85
Upper	0.89	0.88	0.91	0.91	0.90	0.88	0.87
INTRINSIC MOTIVATION							
Estimate	0.84	0.87	0.86	0.89	0.87	0.87	0.87
Lower	0.79	0.84	0.81	0.86	0.83	0.86	0.86
Upper	0.90	0.89	0.91	0.91	0.91	0.89	0.88

INSTRUMENT	*SCHOOL_A*	*SCHOOL_B*	*SCHOOL_C*	*SCHOOL_D*	*SCHOOL_E*	*SCHOOL_F*	TOTAL
Sample Size	81	210	63	200	113	410	1077
GRADE MOTIVATION							
Estimate	0.80	0.80	0.77	0.84	0.78	0.78	0.80
Lower	0.74	0.75	0.68	0.80	0.72	0.74	0.78
Upper	0.87	0.84	0.86	0.87	0.85	0.81	0.82
CAREER MOTIVATION							
Estimate	0.93	0.93	0.87	0.94	0.93	0.91	0.92
Lower	0.90	0.91	0.83	0.92	0.91	0.90	0.92
Upper	0.95	0.94	0.92	0.95	0.95	0.92	0.93

Table 3.7. Standardized factor loadings and fit indices for the confirmatory factor analysis of the measurement model for each instrument by school and all schools combined

INSTRUMENT	*SCHOOL_A*	*SCHOOL_B*	*SCHOOL_C*	*SCHOOL_D*	*SCHOOL_E*	*SCHOOL_F*	TOTAL
Sample Size	81	210	63	200	113	410	1077
Pre-CANS							
Natural Selection	0.224	0.201	0.044	0.229	0.308	0.295	0.305
(range)	0.719	0.745	0.816	0.711	0.838	0.642	0.700
χ^2	278.61	373.53	274.06	426.77	332.12	525.03	1154.57
df	170	170	170	170	170	170	170
χ^2/df	1.639	2.197	1.612	2.510	1.953	3.088	6.792
p-value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
CFI	0.768	0.792	0.704	0.761	0.796	0.785	0.807
RMSEA	0.089	0.076	0.099	0.087	0.092	0.071	0.073
SRMR	0.087	0.076	0.104	0.079	0.081	0.067	0.063
Post-CANS							
Natural Selection	0.195	0.245	0.156	0.248	0.051	0.184	0.256
(range)	0.672	0.746	0.778	0.703	0.840	0.742	0.740
χ^2	325.04	344.84	234.62	447.91	432.17	680.09	1417.60
df	170	170	170	170	170	170	170
χ^2/df	1.912	2.028	1.380	2.634	2.542	4.000	8.339
p-value	<0.001	<0.001	0.001	<0.001	<0.001	<0.001	<0.001
CFI	0.684	0.825	0.752	0.733	0.739	0.739	0.777
RMSEA	0.106	0.070	0.078	0.090	0.117	0.086	0.083
SRMR	0.100	0.068	0.097	0.081	0.097	0.075	0.065
MSU-FORT							
Formal Reasoning	-	-0.233	-	-0.356	-0.101	-0.468	-0.265
(range)	-	0.555	-	0.559	0.516	0.525	0.514
χ^2	-	249.86	-	272.42	266.61	364.71	583.21
df	-	209	-	209	209	209	209
χ^2/df	-	1.196	-	1.303	1.275	1.745	2.790

INSTRUMENT	*SCHOOL_A*	*SCHOOL_B*	*SCHOOL_C*	*SCHOOL_D*	*SCHOOL_E*	*SCHOOL_F*	TOTAL
Sample Size	81	210	63	200	113	410	1077
p-value	-	0.028	-	0.002	0.004	<0.001	<0.001
CFI	-	0.845	-	0.843	0.602	0.818	0.809
RMSEA	-	0.031	-	0.039	0.049	0.043	0.041
SRMR	-	0.062	-	0.062	0.086	0.054	0.044
SMQ-II							
Self-Efficacy	0.537	0.526	0.636	0.630	0.505	0.473	0.549
(range)	0.851	0.822	0.748	0.823	0.768	0.805	0.804
Self-Determination	0.608	0.666	0.545	0.764	0.701	0.669	0.709
(range)	0.834	0.769	0.869	0.814	0.784	0.773	0.787
Intrinsic Motivation	0.547	0.701	0.623	0.678	0.666	0.663	0.658
(range)	0.829	0.829	0.877	0.870	0.871	0.849	0.854
Grade Motivation	0.529	0.511	0.356	0.595	0.366	0.475	0.514
(range)	0.824	0.787	0.905	0.900	0.806	0.811	0.808
Career Motivation	0.793	0.813	0.645	0.820	0.783	0.777	0.805
(range)	0.923	0.883	0.869	0.917	0.884	0.877	0.893
χ^2	531.03	536.33	414.70	609.89	462.31	775.96	1430.77
df	265	265	265	265	265	265	265
χ^2/df	2.004	2.024	1.565	2.301	1.745	2.928	5.399
p-value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
CFI	0.789	0.907	0.835	0.903	0.882	0.911	0.925
RMSEA	0.111	0.070	0.095	0.081	0.081	0.069	0.064
SRMR	0.099	0.061	0.093	0.060	0.073	0.057	0.048

Note: Range of standardized factor loadings and four measures of model fit reported (χ^2, *df*, χ^2/df, *p*-value [two-tailed] – chi-square, degrees of freedom, normed chi-square, and probability distribution; CFI – comparative fit index; RMSEA - root mean square error of approximation; SRMR - standardized root mean square residual).

Building Models of Conceptual Change

For this study, the extended linear regression analytical technique of hierarchical linear modeling (HLM) is used to investigate the relationship among students' post-instruction score as the dependent (response) variable and students' pre-instruction score, formal reasoning ability score, and academic motivation ratings as independent (explanatory) variables, accounting for student demographic variables and controlling for teaching practice and teacher experience at each postsecondary institution. There are four basic assumptions for multiple linear regression analyses. One, the observations are independent of each other; that is, one measurement does not influence another. Two, there is a linear relationship between the dependent variable and each independent variable. Three, the distribution of the response variable conditional on the explanatory variables is assumed to be normally distributed. Put another way, the residuals from the linear model follow a normal distribution. Four, the variance of the residuals is constant for all observations (Ramsey & Schafer, 2013). For the nested data of students in schools, the assumption of independent observations is violated because student observations within the same school or cluster is expected to be more alike than student observations in different schools. HLM accounts for this violation by including random effects that induces correlation between all observations in the same cluster yet assumes no correlation between observations in different clusters. In this study, a random intercept model is used that determines the intra-class correlation (*ICC*) between two student post-instruction scores taken in the same school. *ICC* tells how much information any two student post-instruction scores in the same school share; or stated in practical terms, the correction necessary for the correlation between any two student post-instruction scores from the same school (Galecki & Burzykowski, 2013; Zuur, Ieno, Wlaker, Saveliev, & Smith, 2009). Figure 3.4 (page 112) describes in detail the full model with response quantitative variable, *POST*, and each explanatory quantitative variable *PRE*, *FORMAL*, *SELF_EFF*, *SELF_DET*, *INTRINSIC*, *GRADE*, and *CAREER* at the student level; plus control categorical variables *f.SEX*, *f.MAJ_BIO*, and *f.RACE* at the student level, in addition to control quantitative variables *CWSEI_TPI* and *Semesters_Taught* at the school level. The random intercept, *SCHOOL*, and standard error, ε, are described and the coefficients, βs, are explained.

The “lme” function from the “nlme” package (Pinheiro, Bates, DebRoy, Sarkar, & R Core Team, 2018) is used to perform the analyses for this study. For the Akaike’s An Information Criterion (AIC) method, the maximum likelihood estimator determines estimates of AIC, with $\text{AIC} = -2\log(l_1) + 2p$, such that l_1 is the maximum candidate likelihood, and p is the number of free parameters that includes the βs and variance parameters, which determines the degrees of freedom. The smaller the AIC estimate, the better the predictive properties of the model based on a balance of model fit (l_1) and model complexity ($2p$). A difference of over 2 AIC units is considered evidence in support of true differences in the models being compared. The difference (delta) between the AIC estimate for the model under consideration and the most optimal model with the lowest AIC estimate, $\text{delta} = \text{AIC} - \min(\text{AIC})$, is reported for ease of model comparison (Akaike, 1974; Burnham & Anderson, 2004; Burnham, Anderson, & Huyvaert, 2011).

$POST_{ij} = \mu_{ij} + SCHOOL_i + \varepsilon_{ij}$, where

$$\begin{aligned}\mu_{ij} &= \beta_0 + \beta_1 PRE_{ij} + \beta_2 FORMAL_{ij} + \beta_3 SELF\_EFF_{ij} + \beta_4 SELF\_DET_{ij} + \beta_5 INTRINSIC_{ij} \\ &+ \beta_6 GRADE_{ij} + \beta_7 CAREER_{ij} + \beta_8 D_{f.SEX=Female_{ij}} + \beta_9 D_{f.MAJBIO=No_{ij}} \\ &+ \beta_{10} D_{f.RACE=Minority_{ij}} + \beta_{11} Semesters\_Taught_i + \beta_{12} CWSEI\_TPI_i\end{aligned}$$

such that
$i = 1, \ldots, 6$, are the labels for the schools,
$j = 1, \ldots, n_j$ are the labels for the students within each school,
$SCHOOL_i \sim N(0, \sigma^2_{SCHOOL})$ independent of $\varepsilon_{ij} \sim N(0, \sigma^2_{\varepsilon})$

PRE_{ij}, $FORMAL_{ij}$, $SELF\_EFF_{ij}$, $SELF\_DET_{ij}$, $INTRINSIC_{ij}$, $GRADE_{ij}$, and $CAREER_{ij}$ are quantitative explanatory variable at the observational level (*level one*),

$Semesters\_Taught_i$ and $CWSEI\_TPI_i$ are quantitative explanatory variable contextual at the school level (*level two*),

$D_{f.SEX=Female_{ij}}$, $D_{f.MAJBIO=No_{ij}}$, and $D_{RACE=Minority_{ij}}$ are categorical explanatory variables at the observational level (*level one*),

$D_{f.SEX=Female_{ij}}$ is 1 for female and 0 otherwise; $D_{f.MAJBIO=No_{ij}}$ is 1 for not biology majors and 0 otherwise; and $D_{f.RACE=Minority_{ij}}$ is 1 for a minority ethnicity that is underrepresented in science and 0 otherwise.

β_0 is the intercept, which is the mean POST score when PRE_{ij}, $FORMAL_{ij}$, $SELF\_EFF_{ij}$, $SELF\_DET_{ij}$, $INTRINSIC_{ij}$, $GRADE_{ij}$, $CAREER_{ij}$, $Semesters\_Taught_i$, and $CWSEI\_TPI_i$ are zero.

β_1, β_2, β_3, β_4, β_5, β_6, β_7, β_{11}, and β_{12} are the respective slope for PRE_{ij}, $FORMAL_{ij}$, $SELF\_EFF_{ij}$, $SELF\_DET_{ij}$, $INTRINSIC_{ij}$, $GRADE_{ij}$, $CAREER_{ij}$, $Semesters\_Taught_i$, and $CWSEI\_TPI_i$. Each slope represents the change in mean POST score for every one unit change in the respective quantitative explanatory variable.

β_8, β_9, and β_{10} are the respective change in mean POST score for females versus males, not a biology major versus biology major, and ethnicity underrepresented in science versus ethnicity not underrepresented in the sciences.

Figure 3.4. Full model with fixed effects, random effect, and coefficients described

Standard diagnostic plots for HLM are useful in assessing key statistical assumptions. Figure 3.5 (page 114) displays three diagnostic plots for the full model: Residuals versus Fitted, Normal QQ-Plot of Residuals, and Normal QQ-Plot of Random Effect for SCHOOL (Fox & Weisberg, 2011; Pinheiro, Bates, DebRoy, Sarkar, & R Core Team, 2018). The residual is the difference between the observed value for the dependent variable and the fitted value from the linear model. The plot of the standardized residuals versus the fitted values assesses linearity between the response and quantitative explanatory variables, constant variance of the residuals, and provides a visual check for potential outliers. The distribution of the data is approximately uniform for all fitted values. Therefore, there are no serious concerns about violation of the assumptions of linearity and constant variance. Also, there are no obvious outliers. The normal QQ-plot of residuals graphs standardized quantiles versus theoretical quantiles expected for a normal distribution of the residuals. The data is plotted along the one-to-one line representing the case when the quantiles are equal. This plot assesses the assumption of normality of the residuals. The distribution of the residual is right-skewed, which is evidence of a violation of the normality assumption. However, a sample size of 1140 is large enough for this not to be much of an issue. The normal QQ-plot of the random effect depicts the relationship between the random intercept quantiles and normal quantiles for the six schools in the study. This plot assesses the assumption of normality of the random intercepts. There is some evidence the random intercepts are somewhat heavy tailed relative to the normal distribution conditional on the fixed effects, but the tails are not severe enough to be a serious concern.

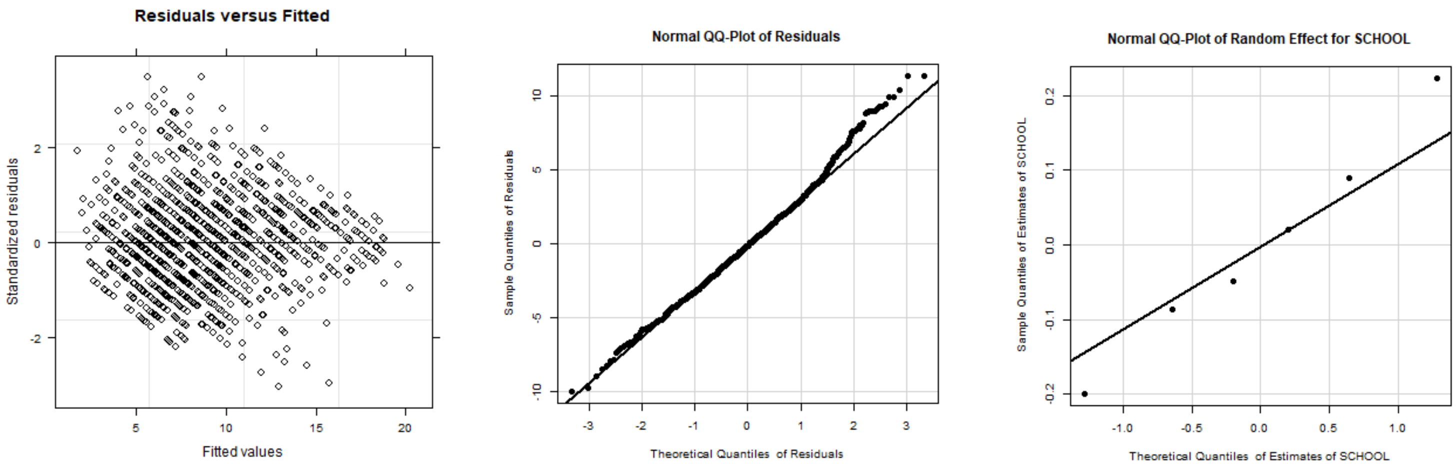


Figure 3.5. Standard diagnostic plots for full model

While the full model delineated in Figure 3.4 is comprehensive and captures important variables that describe change in conceptual understanding of evolution by natural selection, it is also of interest to determine how well, if at all, simpler models explain conceptual change. In addition to the full model (*Model 9*), eight other models examine the relationship of *POST* conditional on *PRE*, accounting for random intercept *SCHOOL*, for the study of conceptual change in this investigation. Table 3.8 (page 116) lists the variables included in each of the nine models. *Model 1* explores the impact of student demographic variables only. *Model 2* looks at just formal reasoning, whereas *Model 3* incorporates formal reasoning and student demographic variables. In addition to student demographic variables, *Model 4* includes formal reasoning and motivation variables related to intrinsic goal orientation. *Model 5* accounts for student demographic variables, as well as examines the role of formal reasoning and motivation variables related to extrinsic goal orientation. *Model 6* comprise all student variables; that is, formal reasoning, intrinsic and extrinsic goal motivation, in addition to student demographic variables. In contrast, *Model 7* consists of all student variables except formal reasoning. *Model 8* investigates only the school level variables of teaching practice and teacher experience. The models summarized in Table 3.8 reflect the investigator's perspective of what variables may be most important in explaining conceptual change, given what is described in the literature. Another approach is to carry out sophisticated statistical analyses to find what are the most parsimonious models for the data. This is discussed in the next section.

Table 3.8. Building mixed effects models to explain post-instruction scores conditional on pre-instruction scores

Variable	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**	**Model 7**	**Model 8**	**Model 9**
Response									
POST	+	+	+	+	+	+	+	+	+
Fixed Effect									
Intercept	+	+	+	+	+	+	+	+	+
Student (level one)									
PRE	+	+	+	+	+	+	+	+	+
FORMAL		+	+	+	+	+			+
SELF_EFF				+		+	+		+
SELF_DET				+		+	+		+
INTRINSIC				+		+	+		+
GRADE					+	+	+		+
CAREER					+	+	+		+
f.SEX	+		+	+	+	+	+		+
f.MAJ_BIO	+		+	+	+	+	+		+
f.RACE	+		+	+	+	+	+		+
School/Instructor (level two)									
Semesters_Taught								+	+
CWSEI_TPI								+	+
Random Effect									
SCHOOL	+	+	+	+	+	+	+	+	+

+ Included

For models of interest, the variance inflation factor (VIF) is determined to check for multicollinearity; that is, any shared information among any explanatory variable and other explanatory variables in the model (Greenwood, 2017). The square root of VIF provides an estimate of how many times larger the standard error would be due to multicollinearity than it would be if the explanatory variables were uncorrelated and independent of each other. The smallest value of VIF is 1, corresponding to no multicollinearity. Acceptable values for VIF are ≤5. Small values of VIF indicate the slope of one variable in the model can be interpret with some trust, holding all other variables constant. In other words, the calculated slopes are stable, and the slope of one variable is not impacted too much by other variables in the model. For severe cases of multicollinearity with VIF of >5 or 10, the variable should be removed from the model as it is redundant in the presence of the other variables with which it shares information that it provides about the response variable (James, Witten, Hastie, & Tibshirani, 2013). The VIFs for the full model are reported in chapter 4.

Parsimonious Models of Conceptual Change

The model selection procedure attempts to find optimal models among a large set of candidate models. The "dredge" function from the "MumIn" package (Bartoń, 2018) is used to fit all possible models starting with the full model. For the Akaike's An Information Criterion (AIC) model selection method, the maximum likelihood estimator determines estimates of AIC. The difference, $\text{delta} = \text{AIC} - \min(\text{AIC})$, is reported for convenient comparison between the candidate model and the most optimal model

(Akaike, 1974; Burnham & Anderson, 2004; Burnham, Anderson, & Huyvaert, 2011). The most parsimonious models that best describe the relationship of *POST* conditional on *PRE*, accounting for random intercept *SCHOOL*, for the study of conceptual change are reported. The VIFs for the final model are also reported (James, Witten, Hastie, & Tibshirani, 2013).

Validity of Study

Several steps were taken to ensure validity of this observational study on what variables impact change in students' conceptual understanding of evolution by natural selection in introductory biology courses across the United States. To achieve internal validity (Mills & Gay, 2016), four specific actions were taken. First, instruments with sound psychometric properties that measure students' understanding of the concept of evolution by natural selection, formal reasoning ability, and academic motivation, in addition to instructors' teaching practice were used in the study. Second, students' demographic variables that may influence conceptual change were included and controlled for in the study. Third, from a list of 258 public four-year colleges and universities in the nation that offer a baccalaureate degree in biology and have 10,000 or more students, six institutions participated in the study. Investigating multiple institutions, as opposed to one or two, and accounting for the teaching experience at each institution strengthens any inferences drawn from the study. Fourth, all the data were collected in the same semester. Also, instructors and students at the six participating institutions followed similar protocols set by the principal investigator's research team in disseminating, completing, and compiling the information gathered from the instruments in this study. To ensure external validity (Mills & Gay, 2016), every attempt was made to obtain a representative sample of introductory biology students. Randomly selected postsecondary institutions were contacted, and the instructors assigned to teach the introductory biology course for majors were invited to participate to avoid obtaining a biased study sample. At the outset, 15 institutions were sought, but only six participated. Every effort was made to comply with human subjects' research protocol at the investigator's institution and the six participating institutions in the study. Students and instructors gave their consent to have their data collected and analyzed. They also gave their permission to publish the data anonymously, as part of an aggregate, to protect their personal as well as institutional identity and confidentiality.

Chapter Summary

Investigating the construct of conceptual change that targets other important constructs besides subject matter knowledge, which is carried out at several institutions is a challenging and difficult feat. It requires strong organization and strategic coordination on the part of the investigators, as well as considerable funding to do it well. Data were collected from six participating postsecondary institutions across the United States on students' post-instruction score as the response variable and students' pre-instruction score, formal reasoning ability score, and academic motivation ratings as explanatory variables, accounting for student demographic variables of biological sex, declaration as a biology major, and representation of racial group in science, plus controlling for teaching practice and experience at each postsecondary institution for the study of concept change for the topic of evolution by natural selection. The psychometric instruments used in the study were reasonably valid and reliable for the sample of introductory biology students. Analyses are carried out on the full model as well as simpler models to determine how well the models under study describe and explain conceptual change.

CHAPTER FOUR

RESULTS

Introduction

This study attempts to answer two research questions regarding which hierarchical linear regression model best explains undergraduate introductory biology students' change in conceptual understanding of evolution by natural selection for students studying at six postsecondary institutions across the United States. Accounting for certain student demographics, in addition to teaching practice and teacher experience, the first question asks, "To what extent do students' formal reasoning ability and academic motivation predict levels of conceptual change?" The second question asks, "Which variable, or combination of variables, that is, formal reasoning ability, academic motivation, teaching practice, teacher experience, and student demographics, are most likely to predict levels of conceptual change?" In the following sections, results pertaining to each of these two questions are presented and discussed, beginning with research question 1 and then followed by research question 2.

Research Question 1

To answer the first research question, nine mixed effect models of conceptual change were built and compared to determine how well formal reasoning and academic motivation predict conceptual change. The models were constructed with post-instruction score on the *CANS* instrument as the response variable and pre-instruction score on the *CANS* as an explanatory variable intended to evaluate conceptual change. Explanatory variables of formal reasoning ability and components of academic motivation (self-efficacy, self-determination, intrinsic motivation, grade motivation, and career motivation) were added to investigate if these variables are associated with conceptual change. Student demographic variables of biological sex, declaration as a biology major, and representation of racial group in science were also added to assess whether these variables were needed to explain conceptual change. Instructor variables of teacher experience and teaching practice were added to determine if controlling for the teaching and learning environment was important for conceptual change. Lastly, a random intercept for the postsecondary institutions was included to account for intercorrelation among students at the same school.

In the following subsections, each model is discussed in turn referencing the results in Tables 4.1 and 4.2 for 1140 introductory biology students at six postsecondary institutions that participated in the study. Table 4.1 (pages 124-125) presents the results of the *F*-test with the null hypothesis that the coefficients of the fixed effects are equal to zero, $H_0 : \beta_1 = \beta_2 = \cdots = \beta_i = 0$, and the alternative hypothesis that at least one coefficient of the fixed effects is not equal to zero, $H_a : \beta_i \neq 0;\ i = 1, 2, \cdots, k$, where k represents the respective fixed effect, accounting for all other fixed effects in the model. Put another way, the alternative hypothesis says at least one of the fixed effects is useful in predicting the post-instruction score. The AIC values for the models are also included in Table 4.1. Table 4.2 (pages 126-127) lists the estimates of the coefficients for the fixed effects and the corresponding *p*-values for the *t*-test with the null hypothesis that the coefficient of each fixed effect is equal to zero, $H_0 : \beta_i = 0$, and alternative hypothesis that the coefficient of each fixed effect is not equal to zero, $H_a : \beta_i \neq 0$, holding all other fixed effects constant in the model. The coefficient of a quantitative fixed effect tells how much the post-instruction score changes for every 1-unit increase in the fixed effect. While, the coefficient of a categorical fixed effect indicates a difference in the post-instruction score for a reference level and another level for the fixed effect. Table 4.2 also reports the random intercept for schools, which is the adjustment needed for the average student's post-instruction score based on fixed effects alone. This determines if there are differences among the average student in each school. Finally, Table 4.2 details the variance for the random intercept related to schools (σ^2_{SCHOOL}) and the variance of the error for the model (σ^2_{ε}). For each model, the *F*-statistic for the most important fixed effects and the associated coefficients for these fixed effects are underscored. In addition, σ^2_{SCHOOL} and σ^2_{ε} are used to calculate the intra-class correlation (*ICC*) for the models. The *ICC* corrects for the intercorrelation among post-instruction scores for students in the same school. The section ends with a discussion on which fixed effects are prevalent across the models, and based on the models' AIC values which of the nine models best predicts levels of conceptual change.

Table 4.1. F-statistic and degrees of freedom for fixed effects for mixed effects models built to explain post-instruction scores conditional on pre-instruction scores with varying student and instructor variables

Variable	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**	**Model 7**	**Model 8**	**Model 9**
Fixed Effect									
Intercept	146.4***	1.997	3.783[a]	0.182	0.325	0.019	6.928**	34.27***	3.285[a]
	1, 1130	1, 1132	1, 1129	1, 1126	1, 1127	1, 1124	1, 1125	1, 1133	1, 1124
Student (level one)									
PRE	951.3***	677.7***	648.0***	617.3***	646.6***	609.2***	868.4***	1051***	609.6***
	1, 1130	1, 1132	1, 1129	1, 1126	1, 1127	1, 1124	1, 1125	1, 1133	1, 1124
FORMAL		133.0***	122.8***	118.3***	120.2***	119.2***			117.6***
		1, 1132	1, 1129	1, 1126	1, 1127	1, 1124			1, 1124
SELF_EFF				0.450		0.793	3.598[a]		1.375
				1, 1126		1, 1124	1, 1125		1, 1124
SELF_DET				0.036		0.009	0.615		0.029
				1, 1126		1, 1124	1, 1125		1, 1124
INTRINSIC				2.439		4.048*	2.343		3.353[a]
				1, 1126		1, 1124	1, 1125		1, 1124
GRADE					0.082	0.025	0.393		0.103
					1, 1127	1, 1124	1, 1125		1, 1124
CAREER					0.135	1.869	0.950		1.600
					1, 1127	1, 1124	1, 1125		1, 1124
f.SEX	0.156		0.095	0.009	0.142	0.029	0.031		0.024
	1, 1130		1, 1129	1, 1126	1, 1127	1, 1124	1, 1125		1, 1124
f.MAJ_BIO	0.002		0.017	0.193	0.001	0.021	0.183		0.242
	1, 1130		1, 1129	1, 1126	1, 1127	1, 1124	1, 1125		1, 1124
f.RACE	14.33***		5.330*	4.660*	5.179*	4.416*	12.84***		5.387*
	1, 1130		1, 1129	1, 1126	1, 1127	1, 1124	1, 1125		1, 1124

Variable	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
School/Instructor (level two)									
Semesters_Taught								16.22*	11.83*
								1, 3	1, 3
CWSEI_TPI								17.61*	15.84*
								1, 3	1, 3
Model Index									
AIC	6080.66	5964.97	5965.56	5965.35	5969.17	5967.22	6079.65	6085.51	5963.34
delta AIC	117.32	1.63	2.22	2.01	5.83	3.88	116.31	122.17	0.00

degrees of freedom listed below *F*-statistic for each fixed effect

[a]$p < 0.1$, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$ (two-tailed, fitted marginally)

Table 4.2. Estimates for mixed effects models built to explain post-instruction scores conditional on pre-instruction scores with varying student and instructor variables

Variable	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
Fixed Effect									
Intercept	4.506***	0.612	0.950[a]	-0.330	0.523	-0.131	2.532**	2.671***	-1.994[a]
Student (level one)									
PRE	0.727***	0.625***	0.620***	0.613***	0.620***	0.610***	0.711***	0.740***	0.608***
FORMAL		0.351***	0.340***	0.336***	0.338***	0.339***			0.333***
SELF_EFF				0.122		0.166	0.371[a]		0.217
SELF_DET				-0.033		0.018	-0.149		0.031
INTRINSIC				0.245		0.353*	0.282		0.321[a]
GRADE					0.057	-0.033	0.139		-0.068
CAREER					0.044	-0.206	-0.155		-0.190
f.SEXFemale	-0.089		-0.066	-0.021	-0.081	0.038	0.041		0.034
f.MAJ_BIONo	0.011		-0.029	0.101	0.007	0.034	0.106		0.115
f.RACEMinority	-1.051***		-0.616*	-0.579*	-0.609*	-0.564*	-1.000***		-0.618*
School/Instructor (level two)									
Semesters_Taught								-0.081*	-0.075*
CWSEI_TPI								5.912*	6.338*
Random Effect									
Intercept									
SCHOOL_A	-0.642	-0.533	-0.436	-0.448	-0.433	-0.478	-0.758	0.000	-0.035
SCHOOL_B	0.015	0.294	0.206	0.188	0.206	0.189	-0.005	0.000	0.135
SCHOOL_C	0.161	0.141	0.151	0.127	0.145	0.144	0.155	0.000	0.068
SCHOOL_D	-0.182	-0.403	-0.357	-0.353	-0.348	-0.367	-0.180	0.000	-0.136
SCHOOL_E	0.866	0.981	0.924	0.935	0.920	0.951	0.939	0.000	0.015
SCHOOL_F	-0.217	-0.480	-0.489	-0.450	-0.489	-0.439	-0.150	0.000	-0.047

Variable	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**	**Model 7**	**Model 8**	**Model 9**
Variance									
σ^2_{SCHOOL}	0.279	0.358	0.311	0.307	0.307	0.317	0.326	0.000	0.026
σ^2_{ε}	11.889	11.826	10.719	10.661	10.716	10.640	11.767	12.058	10.643

[a]$p < 0.1$, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$ (two-tailed, fitted marginally)

Model 1: Student Demographics

The goal of Model 1 is to estimate the relationship between student demographic variables (biological sex, declaration as a biology major, and representation of racial group in science) and levels of conceptual change. Specifically, Model 1 is the baseline model used to assess, holding pre-instruction score constant, are student demographics needed to explain post-instruction score. The results show representation of racial group in science is the only important student demographic that explains post-instruction score for the sample of introductory biology students in the study (F (1, 1130) = 14.33, two-sided p = 0.0002), holding pre-instruction score constant, and controlling for the other student demographics in the model. Students who identify with a racial group in the minority in science (not White or Asian) have a mean post-instruction score of 1.051 points lower than students who identify with a racial group in the majority in science (White or Asian). Accounting for pre-instruction score, biological sex, declaration as a biology major, and representation of racial group in science, the predicted post-instruction score for the average student in *SCHOOL_A*, *SCHOOL_D*, and *SCHOOL_F* is 0.642, 0.182, and 0.217 points less, respectively, than the post-instruction score predicted by the fixed effects alone. However, the predicted post-instruction scores for the average student in *SCHOOL_B*, *SCHOOL_C*, and *SCHOOL_E*, respectively, is greater by 0.015, 0.161, and 0.866 points. These findings suggest there are differences in levels of conceptual change among the average student in one participating school compared to another, such that levels of conceptual change are higher in three schools and lower in the other three schools, controlling for pre-instruction score and student demographics. The intra-class correlation (*ICC*) is calculated to determine the correction necessary due to dependence of post-instruction scores among students in the same school. For Model 1, across all schools, $ICC = \hat{\sigma}^2_{SCHOOL} / (\hat{\sigma}^2_{SCHOOL} + \hat{\sigma}^2_{\varepsilon}) = 0.279/(0.279 + 11.889) = 0.023$, which is less than 0.01% of the maximum post-instruction score of 20 points. This represents only a minimal correction to the post-instruction score due to clustering of students in the same school. Next, ignoring student demographics, the role of formal reasoning ability is investigated.

<u>Model 2: Formal Reasoning</u>

Model 2 is the simplest model tested with only formal reasoning ability as an explanatory variable for levels of conceptual change. The findings indicate there is very strong evidence formal reasoning ability is a significant predictor of post-instruction score (F (1, 1132) = 133.0, two-sided $p < 0.0001$), holding pre-instruction score constant. For every additional question correct on the *MSU-FORT* instrument, the mean post-instruction score increases by 0.351 points. The predicted post-instruction scores for the average student in *SCHOOL_A*, *SCHOOL_D*, and *SCHOOL_F*, respectively, are 0.533, 0.403, and 0.480 points less than the predicted post-instruction score with pre-instruction score and formal reasoning ability as sole predictors. Correspondingly, the predicted post-instruction scores are 0.294, 0.141, and 0.981 points greater for the average student in *SCHOOL_B*, *SCHOOL_C*, and *SCHOOL_E*. That is, on average, students in three schools have lower levels of conceptual change and students in the other three schools have higher levels of conceptual change, controlling for pre-instruction score and formal reasoning ability. The *ICC* for Model 2 is 0.029. Only a slight correction to the post-instruction score is needed due to intercorrelation among students in the same school. The following model explores, given formal reasoning ability as an explanatory variable, is it important to also account for student demographics?

Model 3: Formal Reasoning and Student Demographics

The purpose of Model 3 is to examine how well formal reasoning ability predicts levels of conceptual change, controlling for student demographics. Model 3 builds on Model 2 to test the extent to which formal reasoning ability predicts post-instruction score, holding pre-instruction score constant, and accounting for student demographics of biological sex, declaration as a biology major, and representation of racial group in science. The analysis indicates there are two important predictors other than pre-instruction score for Model 3. There is very strong evidence formal reasoning ability is a principal predictor (F (1, 1129) = 122.8, two-sided p < 0.0001), and moderate evidence representation of racial group in science is a good predictor (F (1, 1129) = 5.330, two-sided p = 0.0211) of post-instruction score, holding pre-instruction score constant, and controlling for the other student demographics in the model. For each added question correct on *MSU-FORT*, on average, the post-instruction scores of students increase by 0.340 points. The mean post-instruction scores of students who identify with a racial group in the minority in science is 0.616 points lower than students who identify with a racial group in the majority in science. The predicted post-instruction scores for the average student in *SCHOOL_A*, *SCHOOL_D*, and *SCHOOL_F*, respectively, are 0.436, 0.357, and 0.489 points less than the predicted post-instruction score, accounting for pre-instruction score, formal reasoning ability, and student demographics. In comparison, the predicted post-instruction scores are 0.206, 0.151, and 0.924 points greater in *SCHOOL_B*, *SCHOOL_C*, and *SCHOOL_E*, respectively. Model 3 has an *ICC* of 0.029, a very small correction for post-instruction score. In the next model, components of academic motivation are considered, along with formal reasoning ability.

<u>Model 4: Formal Reasoning, Intrinsic Goal, and Student Demographics</u>

The objective of Model 4 is to determine the relationship among formal reasoning ability, intrinsic goal orientation variables, and levels of conceptual change, controlling for student demographics. Model 4 extends Model 3 with the inclusion of intrinsic goal orientation variables of self-efficacy, self-determination, and intrinsic motivation. There are two important predictors other than pre-instruction score for Model 4. There is very strong evidence formal reasoning ability is a primary predictor (F (1, 1126) = 118.3, two-sided $p < 0.0001$), and moderate evidence representation of racial group in science is a good predictor (F (1, 1126) = 4.660, two-sided $p = 0.0311$) of post-instruction score, holding pre-instruction score constant, and accounting for the other variables in the model. For every additional question correct on *MSU-FORT*, the mean post-instruction scores of students increase by 0.336 points. On average, the post-instruction scores of students who identify with a racial group in the minority in science is 0.579 points lower than students who identify with a racial group in the majority in science. The predicted post-instruction scores for the average student in *SCHOOL_A*, *SCHOOL_D*, and *SCHOOL_F*, respectively, are 0.448, 0.353, and 0.450 points less than the predicted post-instruction score, accounting for pre-instruction score, formal reasoning ability, intrinsic goal orientation, and student demographics. In comparison, the predicted post-instruction scores are 0.188, 0.127, and 0.935 points greater in *SCHOOL_B*, *SCHOOL_C*, and *SCHOOL_E*, respectively. The *ICC* for Model 4 is 0.028, a negligible correction for post-instruction score. In the following model, formal reasoning ability and extrinsic goal orientation are considered.

Model 5: Formal Reasoning, Extrinsic Goal, and Student Demographics

The goal of Model 5 is to investigate the relationship of formal reasoning ability and extrinsic goal orientation variables on levels of conceptual change, controlling for student demographics. Model 5 includes extrinsic goal orientation variables of grade motivation and career motivation. Like Model 4, the results for Model 5 show there are two important predictors other than pre-instruction score. There is very strong evidence formal reasoning ability is a principal predictor (F (1, 1127) = 120.2, two-sided $p <$ 0.0001), and moderate evidence representation of racial group in science is a good predictor (F (1, 1127) = 5.179, two-sided p = 0.0230) of post-instruction score, holding pre-instruction score constant, and accounting for the other variables in the model. For each added question correct on *MSU-FORT*, on average, the post-instruction scores of students increase by 0.338 points. The mean post-instruction scores of students who identify with a racial group in the minority in science is 0.609 points lower than students who identify with a racial group in the majority in science. The predicted post-instruction scores for the average student in *SCHOOL_A*, *SCHOOL_D*, and *SCHOOL_F*, respectively, are 0.433, 0.348, and 0.489 points less than the predicted post-instruction score, accounting for pre-instruction score, formal reasoning ability, extrinsic goal orientation, and student demographics. Correspondingly, the predicted post-instruction scores are 0.206, 0.145, and 0.920 points greater for the average student in *SCHOOL_B*, *SCHOOL_C*, and *SCHOOL_E*. Model 5 has an *ICC* of 0.028, a minimal correction for post-instruction score. The next model incorporates intrinsic and extrinsic goal orientation variables, in addition to formal reasoning ability.

Model 6: Formal Reasoning, Intrinsic Goal, Extrinsic Goal, and Student Demographics

The purpose of Model 6 is to estimate how well formal reasoning ability and academic motivation (intrinsic and extrinsic goal orientations) predict levels of conceptual change, controlling for student demographics. There are three important predictors other than pre-instruction score for Model 6. There is very strong evidence formal reasoning ability is a key predictor (F (1, 1124) = 119.2, two-sided $p < 0.0001$), as well as moderate evidence intrinsic motivation (F (1, 1124) = 4.048, two-sided p = 0.0445) and representation of racial group in science (F (1, 1124) = 4.416, two-sided p = 0.0350) are good predictors of post-instruction score, holding pre-instruction score constant, and accounting for the other variables in the model. For every additional question correct on *MSU-FORT*, the mean post-instruction scores of students increase by 0.339 points. For intrinsic motivation, a 1-unit increase on a 5-point Likert scale corresponds to a 0.353-point increase in mean post-instruction score. Students who identify with a racial group in the minority in science have a mean post-instruction score that is 0.564 points lower than students who identify with a racial group in the majority in science. The predicted post-instruction scores for the average student in *SCHOOL_A*, *SCHOOL_D*, and *SCHOOL_F*, respectively, are 0.478, 0.367, and 0.439 points less than the predicted post-instruction score, accounting for pre-instruction score, formal reasoning ability, academic motivation, and student demographics. In comparison, the predicted post-instruction scores are 0.189, 0.144, and 0.951 points greater in *SCHOOL_B*, *SCHOOL_C*, and *SCHOOL_E*, respectively. The *ICC* for Model 6 is 0.029, a slight correction for post-instruction score. The next model differs from Model 6 in that it excludes formal reasoning ability as an explanatory variable.

Model 7: Intrinsic Goal, Extrinsic Goal, and Student Demographics

The objective of Model 7 is to determine how well components of academic motivation predict levels of conceptual change, controlling for student demographics. Other than pre-instruction score, there are two important predictors for Model 7. There is modest evidence self-efficacy is a good predictor (F (1, 1125) = 3.598, two-sided p = 0.0581) and strong evidence representation of racial group in science is a significant predictor (F (1, 1125) = 12.84, two-sided p = 0.0004) of post-instruction score, holding pre-instruction score, and accounting for the other variables in the model. For self-efficacy, a 1-unit increase on a 5-point Likert scale corresponds to a 0.371-point increase in the mean post-instruction score. The mean post-instruction scores of students who identify with a racial group in the minority in science is 1.000 points lower than students who identify with a racial group in the majority in science. The predicted post-instruction scores for the average student in *SCHOOL_A*, *SCHOOL_B*, *SCHOOL_D*, and *SCHOOL_F*, respectively, are 0.758, 0.005, 0.180, and 0.150 points less than the predicted post-instruction score, accounting for pre-instruction score, academic motivation, and student demographics. In comparison, the predicted post-instruction scores are 0.155 and 0.939 points greater in *SCHOOL_C* and *SCHOOL_E*, respectively.

The random intercepts for the schools in Model 7 do not completely follow the trend reported for previous models. Excluding formal reasoning ability, and accounting for academic motivation and student demographics only, on average, levels of conceptual change are higher in two schools and lower in the other four schools. This contrasts with the even split among the schools for other models. The *ICC* for Model 7 is 0.027, a very small correction for post-instruction score.

Model 8: Teaching and Learning Environment

The goal of Model 8 is to determine whether controlling for the learning and teaching environment alone explains levels of conceptual change among students. In Model 8, instructors' self-reported measures of their research-based teaching practice and the number of semesters they have taught the introductory biology course were used as predictors of students' post-instruction scores, accounting for pre-instruction scores. The results show both instructor variables are important in Model 8. There is moderate evidence the number of semesters instructors have taught the course (F (1, 3) = 16.22, two-sided p = 0.0275) and the proportion of research-based practice incorporated in instruction (F (1, 3) = 17.61, two-sided p = 0.0247) are good predictors of post-instruction score, holding pre-instruction score constant. For the six instructors in the study, for every additional semester an instructor teaches the course, students' mean post-instruction score decreases by 0.081 points. As the proportion of researched-based teaching practice incorporated in instruction increases from not incorporated (0.00) to fully incorporated (1.00), students' mean post-instruction score increases by 5.912 points. This represents a large change in post-instruction score given that a unit increase reflects the full scale of no research-based instruction to full research-based instruction. The random intercepts for all schools are 0.000 for Model 8. This means there is no discernable difference among the average student in all six participating schools, accounting only for teacher experience and teaching practice. The *ICC* for Model 8 is 0.000. Therefore, there are no corrections necessary for students' post-instruction score as a result of intercorrelation among students in the same school. The following model includes all the explanatory variables previously discussed.

Model 9: Full Model (All Variables)

The purpose of Model 9 is to test whether the full model with all the explanatory variables is needed to estimate levels of conceptual change. Model 9 comprise response variable, post-instruction score, and fixed effects pre-instruction score, formal reasoning ability, components of academic motivation, student demographics, as well as teaching practice and teacher experience. Other than pre-instruction score, there are five important predictors for Model 9. There is very strong evidence formal reasoning ability is a primary predictor (F (1, 1124) = 117.6, two-sided $p < 0.0001$), plus modest evidence intrinsic motivation (F (1, 1124) = 3.353, two-sided $p = 0.0674$), and moderate evidence representation of racial group in science (F (1, 1124) = 5.387, two-sided $p = 0.0205$) are good predictors of post-instruction score, holding pre-instruction score constant, and accounting for the other variables in the model. There is also moderate evidence teacher experience (F (1, 3) = 11.83, two-sided $p = 0.0413$) and teaching practice (F (1, 3) = 15.84, two-sided $p = 0.0284$) are good predictors of post-instruction score, holding pre-instruction score constant, and accounting for the other variables in the model.

Table 4.3 (page 141) details the results for the full model. Estimates of the fixed effects, random intercept and its variance, in addition to the error variance, the standard error, *t*-statistic, *p*-values, and variance inflation factor (VIF) are reported. For every additional question correct on *MSU-FORT*, the mean post-instruction scores of students increase by 0.333 points. For intrinsic motivation, a 1-unit increase on a 5-point Likert scale corresponds to a 0.321-point increase in mean post-instruction score. Students who identify with a racial group in the minority in science have a mean post-instruction score that is 0.618 points lower than students who identify with a racial group in the majority in science. For every additional semester an instructor teaches the introductory biology course, students' mean post-instruction score decreases by 0.075 points. As the proportion of researched-based teaching practice increases from nonexistent to ideal, students' mean post-instruction score increases by 6.338 points. The predicted post-instruction scores for the average student in *SCHOOL_A*, *SCHOOL_D*, and *SCHOOL_F*, respectively, are 0.035, 0.136, and 0.047 points less than the predicted post-instruction score, accounting for pre-instruction score, formal reasoning ability, academic motivation, student demographics, teacher experience, and teaching practice. In comparison, the predicted post-instruction scores are 0.135, 0.068, and 0.015 points greater in *SCHOOL_B*, *SCHOOL_C*, and *SCHOOL_E*, respectively. The *ICC* for Model 9 is 0.002, a negligible correction for post-instruction score. Lastly, for the full model, there is no evidence multicollinearity is a serious concern as the VIFs for the fixed effects at the student and school levels, respectively, are all ≤3.5. A discussion follows on the fixed effects that are most important among the nine mixed effects models and which of the nine model best explain levels of conceptual change.

Comparison of the Nine Mixed Effects Models of Conceptual Change

Across the nine mixed effect models, certain fixed effects seem to be important in predicting levels of conceptual change; that is, post-instruction score, holding pre-instruction score constant. Pre-instruction score, formal reasoning ability, and representation of racial group in science are fixed effects that are significant predictors in all models in which they are included. Intrinsic motivation is an important predictor in two of four models in which it appears. While, teacher experience and teaching practice are good predictors in the two models that control for the teaching and learning environment. For all nine models, every additional pre-instruction question that is correct on the *CANS*, students' mean post-instruction score increases by 0.608 to 0.740 points. For models that include formal reasoning ability, each question that is correct on the *MSU-FORT* instrument results in a 0.333 to 0.351-point increase in students' mean post-instruction score. The difference between the mean post-instruction score for students who identify with a racial group in the minority in science and those who identify with a racial group in the majority ranges from 0.564 to 1.051 points, with the minority group being repeatedly lower. For the six instructors in the study, an increase in the number of semesters that an instructor teaches the introductory biology course results in a decrease in students' mean post-instruction of 0.075 to 0.081 points. However, an increase in research-based teaching practice in instruction corresponds to a 5.912 to 6.338-point increase in students' mean post-instruction score. The random intercepts for the participating schools follow a similar trend for seven of the nine models. Consistently, three schools have higher post-instruction scores (0.015 to 0.981 points) and three schools have lowers post-instruction scores (0.035 to 0.642 points) than the post-instruction score predicted by the fixed effects alone. Only Model 7 that did not include formal reasoning ability as a fixed effect has a slightly different trend and Model 8 that did not control for any student fixed effects has a very different trend.

In comparing the predictive power of mixed effects models, one statistic that is quite useful in ranking models is the AIC value. The lower the AIC value, the better the model is at reproducing the data for the response variable, given the explanatory variables in the model. In other words, the more optimal is the model at predicting the response variable. Based on the analyses of the nine mixed effects models, the most optimal model in predicting post-instruction score, conditional on pre-instruction score, is the full model, Model 9, with 5963.34 AIC units. Therefore, it is prudent to compare the eight other models to Model 9, and where instructive, the other models to each other. Model 1 controls for student demographics. Model 1 with 6080.66 AIC units is 117.32 AIC units worse than Model 9. Model 2 accounts only for formal reasoning ability. Model 2 is 5964.97 AIC units, and it is 1.63 AIC units worse than Model 9. Model 3 builds on Model 2 by including formal reasoning ability, in addition to student demographics. Model 3 is 5965.56AIC units, which is 2.22 AIC units worse than Model 9 and 0.59 AIC units worse than Model 2. Model 4 extends Model 3 by incorporating intrinsic goal orientation as an added fixed effect. Model 4, with 5965.35 AIC units, is 0.21 AIC units better than Model 3 and 2.01 AIC units worse than Model 9. Model 5 similarly builds on

Model 3, but includes extrinsic goal orientation, in addition to formal reasoning ability and student demographics. The AIC unit for Model 5 is 5969.17, which is 3.61 AIC units worse than Model 3, 3.82 AIC units worse than Model 4, and 5.83 AIC units worse than Model 9. Model 6 accounts for all the student variables, formal reasoning ability, intrinsic and extrinsic goal orientations, as well as student demographics. The AIC unit for Model 6 is 5967.22, which is 2.25 AIC units worse than Model 2, 1.66 AIC units worse than Model 3, 1.07 AIC units worse than Model 4, and 3.88 AIC units worse than Model 9. Model 7 differs from Model 6 in that it does not include formal reasoning ability, only academic motivation, and student demographics. The AIC unit for Model 7 is 6079.65, which is 112.43 AIC units worse than Model 6 and 116.31 AIC units worse than Model 9. Model 8 disregards student variables and only controls for teaching practice and teacher experience. The AIC unit for Model 8 is 6085.51, which is 4.85 AIC units worse than Model 1 that only controls for student demographics, and 122.17 AIC units worse than Model 9. The full model contains fixed effects that do not seem to be important in predicting post-instruction score, conditional on pre-instruction score. In the following section, the full model is used to find more parsimonious models that predict levels of conceptual change.

Table 4.3. Estimates and statistics for full mixed effects model explaining post-instruction scores conditional on pre-instruction

Fixed Effect						
Variables	Estimate	*SE*	*df*	t^{+}	*p*	VIF
Intercept	-1.994	1.101	1124	-1.812	0.07	-
Student (level one)						
PRE	0.608	0.025	1124	24.69	<0.0001	1.096
FORMAL	0.333	0.031	1124	10.845	<0.0001	1.967
SELF_EFF	0.217	0.185	1124	1.172	0.2413	1.878
SELF_DET	0.031	0.181	1124	0.169	0.8660	2.312
INTRINSIC	0.321	0.175	1124	1.831	0.0674	1.477
GRADE	-0.068	0.212	1124	-0.321	0.7484	2.191
CAREER	-0.190	0.151	1124	-1.265	0.2062	1.175
f.SEXFemale	0.034	0.222	1124	0.154	0.8779	1.287
f.MAJ_BIONo	0.115	0.233	1124	0.492	0.6231	1.065
f.RACEMinority	-0.618	0.266	1124	-2.321	0.0205	1.096
School/Instructor (level two)						
Semesters_Taught	-0.075	0.022	3	-3.439	0.0413	3.414
CWSEI_TPI	6.338	1.593	3	3.979	0.0284	3.468
Random Effect						
Intercept						
SCHOOL_A	-0.035					
SCHOOL_B	0.135					
SCHOOL_C	0.068					
SCHOOL_D	-0.136					
SCHOOL_E	0.015					
SCHOOL_F	-0.047					
Variance						
σ^2_{SCHOOL}	0.026					
σ^2_{ε}	10.643					

SE – Standard Error
df – degrees of freedom
$^{+}$*t*-statistic fitted marginally
p – probability distribution (two-tailed)
VIF - Variance Inflation Factor

Research Question 2

To answer the second research question, a hierarchical linear regression analysis ("MumIn" package, Bartoń, 2018) is performed that operates under the hypothesis that the full model is the best complex model that explains levels of conceptual change. The "MumIn" algorithm determines if there are simpler models that consists of one or more fixed effects contained in the full model that better predict post-instruction score, holding pre-instruction score constant, for the 1140 introductory biology students at the six postsecondary institutions in the study. Only significant predictors are reported by this analytical procedure. Table 4.4 (page 144) summarizes the results for the six most parsimonious models of conceptual change that describe introductory biology students' understanding of evolution by natural selection. The β coefficients for the quantitative variables and an indicator ("+") for the categorical variables contained in the six top models obtained from the full model are tabulated.

Parsimonious Models

All six top parsimonious models reported in Table 4.4 are within 1.38 AIC units of each other. This means that statistically speaking, there is no real difference among the six models since they differ by less than 2 AIC units. Therefore, it is more useful to discuss which fixed effects predominate in the top six models to get a better sense of the variables that are most important in predicting levels of conceptual change. Controlling for pre-instruction score and the other variables in the respective models, of the student variables, formal reasoning ability is the only variable that appears in all six models with a positive coefficient that ranges from 0.328 to 0.337. This translates to for every additional question correct on the *MSU-FORT* instrument, students' mean post-instruction score increases by about 0.3 out of a maximum of 20 points. Intrinsic motivation is in five of the six models with coefficients ranging from 0.189 to 0.408. This corresponds to a 1-unit increase on a 5-point Likert scale results in roughly a 0.2 to 0.4-point increase in mean post-instruction scores for students. Self-efficacy shows up in three of six models with coefficients from 0.165 to 0.292, indicating a 1-unit increase on a 5-point Likert scale means about a 0.2 to 0.3-point increase in the mean post-instruction score. There is a negative coefficient for career motivation in two of the six models with an approximate 0.2 decrease in mean post-instruction score for a 1-unit increase on a 5-point Likert scale. Representation of racial group in science is present in all six models, but declaration as a biology major is present in only one model. Both teacher experience and teaching practice appear in all six models. Generally, for the six instructors in the study, their students' mean post-instruction score decreases by less than 0.1 point for every additional semester an instructor teaches the course. In contrast, on average, students' post-instruction score increases by about 6 points as instructors move from one extreme of incorporating no research-based teaching practices to the other extreme of fully integrating best teaching practices in their classrooms. The consistency of the fixed effects across the six models and the small differences in the AIC values suggest the most parsimonious model is a good choice for the final model of conceptual change for evolution by natural selection. This final model is discussed next.

Table 4.4. Top six mixed effects models explaining post-instruction scores conditional on pre-instruction scores

Variable	**Model I**	**Model II**	**Model III**	**Model IV**	**Model V**	**Model VI**
Fixed Effect						
Intercept	-1.776	-1.591	-1.925	-1.901	-2.064	-2.069
Student (level one)						
PRE	0.614	0.611	0.607	0.612	0.611	0.614
FORMAL	0.335	0.337	0.333	0.328	0.331	0.334
SELF_EFF			0.209	0.292	0.165	
SELF_DET						
INTRINSIC	0.271	0.408	0.325		0.189	0.294
GRADE						
CAREER		-0.181	-0.210			
f.SEXFemale						
f.MAJ_BIONo						+
f.RACEMinority	+	+	+	+	+	+
School/Instructor (level two)						
Semesters_Taught	-0.071	-0.072	-0.074	-0.074	-0.073	-0.071
CWSEI_TPI	5.990	6.043	6.195	6.301	6.107	6.211
Random Effect						
SCHOOL	+	+	+	+	+	+
Model Index						
df	9	10	11	9	10	10
Log Likelihood	-2968.50	-2967.61	-2966.85	-2968.87	-2968.02	-2968.19
AIC	5955.0	5955.2	5955.7	5955.7	5956.0	5956.4
delta AIC	0.00	0.21	0.69	0.74	1.03	1.38

+ Included

<u>Final Model</u>

The final model of conceptual change for evolution by natural selection is the optimal mixed effect model with response post-instruction score and pre-instruction score, formal reasoning ability, intrinsic motivation, representation of racial group in science, teacher experience, and teaching practice as fixed effects, accounting for the random intercepts of schools. It is the candidate model with the lowest AIC value generated from the "MumIn" package (see Table 4.4). The optimal model suggests, given model fit and model complexity, the combination of explanatory variables in the model most likely reproduces the data for the post-instruction score of the 1140 introductory biology students at the six postsecondary institutions in the study. The three standard diagnostic plots for the final model (Figure 4.1 on page 147) indicate there are no real concerns for the assumptions of linearity, constant variance, normality of residuals given the sample size, and normality of the random intercepts. Plus, there are no clear outliers in the data. Table 4.5 (page 148) reports the estimate and the 95% confidence interval (95% CI) for the coefficients, in addition to the VIFs, for the fixed effects. The random intercepts for the schools, the variance for the intercepts, and the variance of the residuals are also reported. For the final model, multicollinearity is not a concern, $1.04 \leq VIF \leq 2.63$. The predicted post-instruction scores for the average student in *SCHOOL_A*, *SCHOOL_D*, and *SCHOOL_F*, respectively, are 0.048, 0.199, and 0.087 points less than the predicted post-instruction score accounting for pre-instruction score, formal reasoning ability, intrinsic motivation, representation of racial group in science, teacher experience, and teaching practice only. Correspondingly, the predicted post-instruction scores are 0.223, 0.090, and 0.021 points greater for the average student in *SCHOOL_B*, *SCHOOL_C*, and *SCHOOL_E*. This indicates there are differences in levels of conceptual change among the average student in one participating school compared to another, such that levels of conceptual change are higher in three schools and lower in the other three schools. The *ICC* is 0.004, which is a negligible correction to the post-instruction scores for students in the same school. This final model with seven fixed effects is 8.34 AIC units better than the full model with thirteen fixed effects in the model.

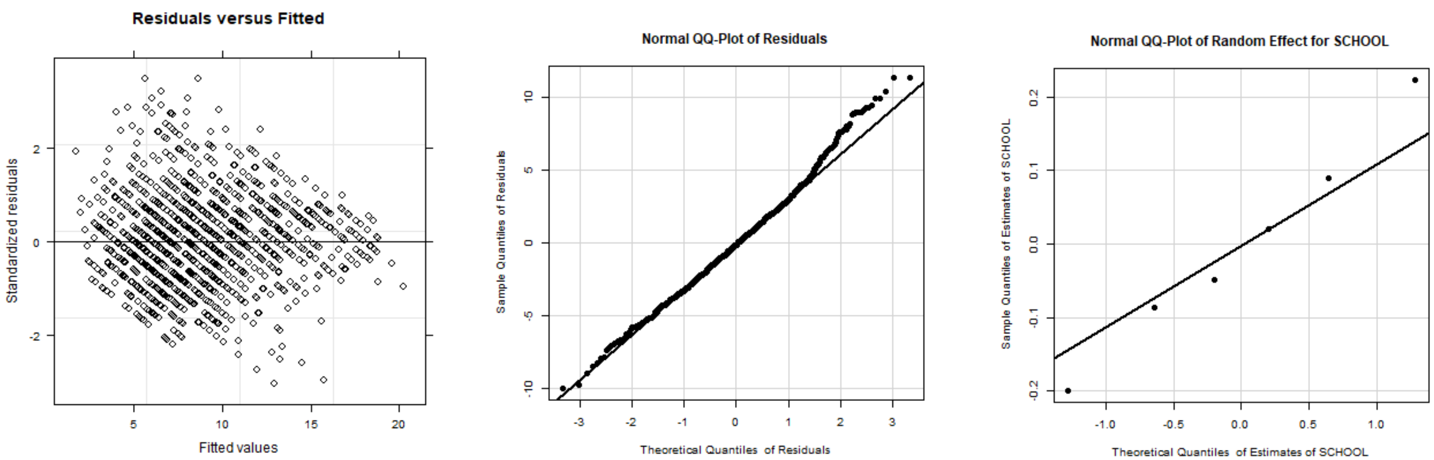


Figure 4.1. Standard diagnostic plots for final model

Table 4.5. Estimates of fixed and random effects with 95% confidence intervals for final model explaining post-instruction scores conditional on pre-instruction scores

Fixed Effect				
Variable	Estimate	Lower	Upper	VIF
Intercept	-1.776	-3.311	-0.240	-
Student (level one)				
PRE	0.614	0.566	0.661	1.24
FORMAL	0.335	0.275	0.394	1.25
INTRINSIC	0.271	0.041	0.501	1.04
f.RACEMinority	-0.624	-1.141	-0.108	1.05
School/Instructor (level two)				
Semesters_Taught	-0.071	-0.144	0.002	2.63
CWSEI_TPI	5.990	0.565	11.415	2.62
Random Effect				
Intercept				
SCHOOL_A	-0.048			
SCHOOL_B	0.223			
SCHOOL_C	0.090			
SCHOOL_D	-0.199			
SCHOOL_E	0.021			
SCHOOL_F	-0.087			
Variance				
σ^2_{SCHOOL}	0.045			
σ^2_{ε}	10.667			

ICC – Intra-class correlation

$ICC = \hat{\sigma}^2_{SCHOOL}/(\hat{\sigma}^2_{SCHOOL} + \hat{\sigma}^2_{\varepsilon}) = 0.045/(0.045 + 10.667) = 0.004$

Figure 4.2 (page 151) displays the fixed effects plots for the final model (Fox, 2003). Each fixed effect plot shows the range of values in the data and the corresponding estimate for the coefficient, depicted by the slope of the line in the graph. The plot also shows the uncertainty for each fixed effect over the range of values in the data, denoted by the shaded grey region in the graph. A discussion follows for each fixed effect, with all other variables in the model held constant, for the introductory biology students at the six postsecondary institutions in the study. The predicted post-instruction scores for students on the *CANS* range from 5.000 to 17.000 points with an increase of 0.614 points for every additional pre-instruction question correct on the *CANS*, (95% CI: 0.566 to 0.661). For formal reasoning ability, the predicted post-instruction scores are from a low of 5.750 to a high of 11.750 points. For each added question correct on the *MSU-FORT*, there is an average increase of 0.335 points for the predicted post-instruction scores, (95% CI: 0.275 to 0.394). The uncertainty of the predicted post-instruction scores for introductory biology students at it relates to their intrinsic motivation on a 5-point Likert scale is quite wide over the entire scale, but particularly for the lower end. Students' predicted post-instruction scores range from 8.175 to 9.200 points. For a 1-unit increase on a 5-point Likert scale, the mean post-instruction score increases by 0.271 points, (95% CI: 0.041 to 0.501). The mean post-instruction score is 8.376 points for students who identify with a racial group in the minority in science compared to 9.000 points for the majority group in science, which means students in the minority group in science has a mean post-instruction score 0.624 points lower than students who identify with a majority racial group (95% CI: -1.141, -0.108). For the six instructors in the study, for each additional semester an instructor teaches the introductory biology course, the mean post-instruction scores for students decrease by 0.071 points, (95% CI: -0.144, 0.002). It is worth noting, however, that the uncertainty in predicted post-instruction scores is greater for more experienced instructors. Students' mean post-instruction score ranges from a high of 9.450 points to a low of 7.300 points for instructors teaching the course for 8 to 38 semesters. The proportion of researched-based teaching practice routinely used by instructors are between 0.33 and 0.70, where 0.00 represents no researched-based practice in instruction and 1.00 represents an ideal situation where researched-based teaching practice is fully integrated in everyday instruction. Moving from one end to the other end of the continuum for how much research-based teaching practice is incorporated in routine instruction, the post-instruction scores for students increase, on average, by 5.990 points, (95% CI: 0.565, 11.415) for instructors who do not integrate best practice at one extreme compared to those who comprehensively do at the other extreme.

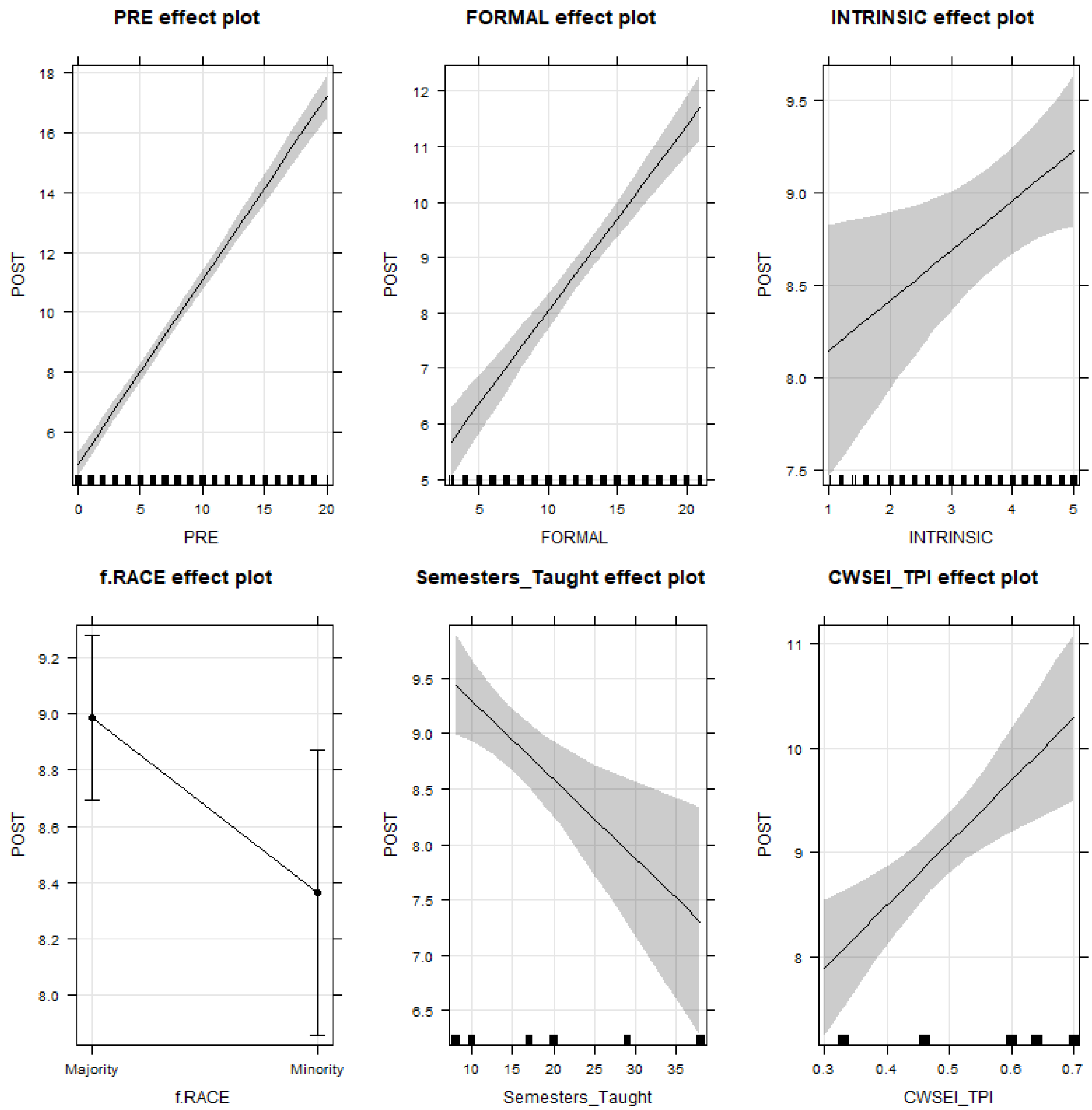


Figure 4.2. Plot of fixed effects explaining post instruction scores for final model

Chapter Summary

This study addresses two research questions. For the first research question regarding to what extent do students' formal reasoning ability and academic motivation predict levels of conceptual change, given a set of control variables for student demographics, and the teaching and learning environment, the full model that includes all variables under consideration in the investigation is the best model. For the second research question of what variable or combination of variables best predict levels of conceptual change, the most parsimonious model is selected as the final model that best operationalizes a model of conceptual change for evolution by natural selection. The final model consists of variables formal reasoning ability, intrinsic motivation, representation of racial group in science, teacher experience, and teaching practice that predict post-instruction score conditional on pre-instruction score.

CHAPTER FIVE

CONCLUSION

Introduction

The present study investigates whether introductory biology students' formal reasoning ability (cognitive construct) and academic motivation (attributional construct) are related to change in conceptual understanding of evolution by natural selection (conceptual change construct), controlling for student demographics, as well as teacher experience and teaching practice (contextual construct). In the science education literature, studies on conceptual change have focused almost exclusively on the manipulation of instruction to bring about a positive change in students' content knowledge and conceptual understanding. This is not at all surprising given the origins of the conceptual change movement in the early 1980s, which emphasized the importance of prior knowledge as a cognitive or rational aspect that is integral to the process of correcting deep-rooted misconceptions (Posner et al., 1982). On a parallel track in the 1970s and 1980s, science education researchers began to explore the role of students' formal reasoning ability as a cognitive construct, beyond prior knowledge, in the learning of science (Lawson, 1985; Lawson & Renner, 1975). A significant breakthrough in conceptual change research occurred in the early 1990s with the publication of a pivotal theoretical model of conceptual change that incorporated extrarational aspects related to students' academic motivation and contextual factors associated with the teaching and learning environment students experience (Pintrich et al., 1993). This extended model of conceptual change brought attention to the importance of extrarational aspects of the conceptual change process. It also underscored the significance of student characteristics such as academic motivation, and in this study, I assert formal reasoning ability to the process of conceptual change. Over the years, not many empirical studies in science education have explored this line of inquiry; that is, what is the relationship between student characteristics and conceptual change in the learning of science (Murphy & Alexander, 2008, Sinatra, 2005).

Only a handful of empirical studies have investigated how formal reasoning ability or academic motivation is related to conceptual change in science. In fact, for the topic of evolutionary biology, only one study in the last ten years has looked at the relationship of self-efficacy and personal interest, two components of academic motivation, and conceptual change (Linnenbrink-Garcia et al., 2012). Moreover, only one study some thirty years ago investigated the association of two student characteristics at once, formal reasoning ability and self-efficacy, on the number of misconceptions retained after instruction on evolution by natural selection (Lawson & Thompson, 1988). Consequently, there is a deficit in the science education literature for empirical studies that examine the relationship between student characteristics and conceptual change. The present study addresses this shortfall in a few ways. One, this study examines the relationship of more than one student characteristic at the same time, formal reasoning ability and five components of academic motivation, and conceptual change. Two, this study accounts for other variables, namely student demographics (biological sex, declaration as a biology major, and representation of racial group in science), teacher experience, and teaching practice, that are likely related to conceptual change to better estimate the relationship of formal reasoning ability, academic motivation, and conceptual change. Three, this study encompasses multiple sites, taking into consideration and controlling for variations among different educational settings to more accurately estimate the relationship of formal reasoning ability, academic motivation, and conceptual change. In the subsequent sections, the significance of the findings from each research question is discussed in the context of what new information is learned and what this knowledge means for conceptual change research moving forward. Implications of the findings for instruction and future practice is presented. Also, concluding remarks on the limitations of the study is offered. Finally, the chapter ends with a brief overview of recommendations for future research.

Research Question 1

The first research question asks to what extent do formal reasoning ability and academic motivation predict levels of conceptual change, controlling for student demographics, plus the teaching and learning environment. The findings are discussed as they relate to each variable examined in the operational models of conceptual change proposed and tested in this investigation.

Formal Reasoning Ability

Consistently, across all models, formal reasoning ability arises as a very important predictor of conceptual change. In fact, formal reasoning ability alone, holding pre-instruction score constant, predicts post-instruction scores almost as well as far more complex models with many variables. This finding suggests that whatever formal reasoning ability measures, whether students' ability to answer scientific questions or a more generic cognitive ability related to learning, formal reasoning ability is a key variable for understanding conceptual change. The classical conceptual change model emphasizes the importance of prior knowledge as a cognitive construct (Posner et al., 1982), and other theoretical models such as the extended conceptual change model that build on it (Pintrich et al., 1993). However, the explicit inclusion of formal reasoning ability as a cognitive construct has been overlooked in theoretical models of conceptual change. Anton Lawson, for many years, was the lone researcher investigating the relationship of formal reasoning ability and science learning. To date, only the research associated with this NSF funded project (#1432577) has explored the relationship of formal reasoning ability and conceptual change for the topic of evolution by natural selection. This study consistently shows that the relationship between formal reasoning ability and conceptual change is quite stable for operational models of conceptual change with formal reasoning ability as the only predictor, and operational models of conceptual change with formal reasoning ability, components of academic motivation, student demographics, and instructor variables as predictors. That is, the change in post-instruction score, holding pre-instruction score constant, is about the same in the simplest model with just formal reasoning ability as the sole predictor and the most complex model with formal reasoning ability and many other variables as predictors. While this study affirms the importance of formal reasoning ability as a principal predictor of conceptual change, it also accentuates the point that the role of formal reasoning ability in conceptual change research has been under appreciated and formal reasoning ability as a cognitive construct has not received the attention it deserves. To understand precisely how formal reasoning ability is related to conceptual change, it is critical to consider if formal reasoning ability is correlated with another variable not currently investigated such as cognitive ability to learn more broadly, not just to carry out scientific investigations. This point will be discussed further in the section on recommendations for future research.

Academic Motivation

The component of academic motivation that is most significantly related to conceptual change depends on whether formal reasoning ability is included or excluded from the operational model of conceptual change. When formal reasoning ability is included in the operational model, intrinsic motivation, which is related to students' personal interest, seems to be the most important component of academic motivation that predicts conceptual change. However, when formal reasoning ability is excluded from the operational model, self-efficacy, which is related to students' confidence in their ability to perform, seems to be most important predictor of conceptual change. This is a very interesting finding that begs the question, what is the relationship among formal reasoning ability, self-efficacy, and intrinsic motivation with regards to conceptual change? Given the respective predictive scales of these three variables, compared to formal reasoning ability, intrinsic motivation and self-efficacy are not as strong predictors of conceptual change. For operational models in which formal reasoning ability and intrinsic motivation are important predictors, an additional question correct for formal reasoning ability and a 1-point increase on a 5-point Likert scale for intrinsic motivation corresponds to about a 1.7% increase in mean post-instruction score. In the case of formal reasoning ability, however, if students were randomly guessing to get each additional question correct, there is only a remote chance we would observe this increase in mean post-instruction score or an increase more extreme. On the other hand, for intrinsic motivation, there is a high probability we would observe this increase in mean post-instruction score or an increase more extreme for a random incremental increase in self-reported personal interest on a 5-point scale. This finding highlights the significance of formal reasoning ability in predicting conceptual change, even when components of academic motivation are included in the operational model. The lack of importance of self-efficacy as a predictor of conceptual change is surprising, given the breadth of research that underscores its importance in learning. This study suggests that the exclusion of a variable related to cognitive ability and the inclusion of only motivation variables in models predicting academic achievement may amount to model misspecification in which the importance of self-efficacy is over stated. This is an argument that has been advanced by other researchers (Valentine, DuBois, & Cooper, 2004). The findings from this study also indicates, in combination with formal reasoning ability, academic motivation related to intrinsic goal orientation is a better predictor of conceptual change than academic motivation related to extrinsic goal orientation. That is, students' personal interest, confidence in their ability to perform well, and control over events related to their learning are more important than a desire for better grades and to prepare for a future career, given their prior knowledge and their reasoning ability to carry out scientific investigations. This is consistent with other findings in the motivation literature (Ryan & Deci, 2000).

Student Demographics

The racial group with which students identify emerges as an important predictor of conceptual change with minority students underperforming when compared to their White or Asian counterparts. The literature suggests that having a role model of a similar race or ethnicity has tangible consequences for student learning (Karunanayake & Nauta, 2004; Perna et al., 2009; Price, 2010), and it seems changing students' conceptual understanding. The importance of race in conceptual change may be explained by the support students perceive they receive or the connection they have with an instructor or mentor of a similar race. As such, students may be more likely to share their vulnerability and lack of understanding if they believe they will be genuinely helped without fear of ridicule or being unfairly dismissed. In a similar vein, students who feel supported by their peers are more likely to benefit from positive relationships that enhance in-class participation, mandatory in-and-out-of-class assignments, out-of-class study groups, and an overall sense of belonging and well-being. Unfortunately, however, minority students often report feelings of isolation and alienation that interfere with their learning, study habits, and ultimately, their academic achievements (Solorzano, Ceja, & Yosso, 2000; Treisman, 1992; Yosso, Smith, Ceja, & Solórzano, 2009). Another possible explanation for the results is many minority students may not have had the academic preparation in STEM that they need to put them in an advantageous position when they take introductory college courses (Chen, 2013; HERI, 2010; Kokkelenberg & Sinha, 2010; Rask, 2010).

Teaching and Learning Environment

There is evidence to suggest that any model of conceptual change must account for the teaching and learning environment students experience. Teaching practice as it relates to the proportion of research-based instruction used seems to be important for bringing about conceptual change. On the other hand, teacher experience appears to have a more nuanced relationship to conceptual change, with more teacher experience corresponding to less favorable conceptual change for the students of the six instructors who participated in this study. It is important to note that the teaching practice explored in this study does not reflect specific practices tailored to induce conceptual change. Regardless, an instructor's disposition to encourage more student discussion and offer opportunities for students to delve deeper and bolster their understanding seems to have a positive association with conceptual change. This is consistent with what is reported in the literature (AAAS, 2011, 2013; Kober, 2015; Singer et al., 2012).

Research Question 2

The second research question asks what variable or combination of variables best predict levels of conceptual change for the variables described in the previous section for the first research question. The final model is proposed as the operational model of conceptual change that optimally captures the important predictors of conceptual change for introductory biology students studying the topic of evolution of natural selection at six postsecondary institutions across the United States.

Operational Model of Conceptual Change

The best model of conceptual change includes variables of formal reasoning ability, intrinsic motivation, representation of racial group in science, teacher experience, and teaching practice. Each of these variables has a unique predictive scale that cannot be compared directly. Nevertheless, the combination of these variables, conditional on pre-instruction score, ably predicts post-instruction scores. The elegance of this operational model of conceptual change cannot be overstated. This final model with three student characteristics is better than the full model with nine student characteristics. This means that student characteristics of formal reasoning ability, intrinsic motivation, and the racial group with which students identify, along with instructor variables of teacher experience and teaching practice do a very good job of predicting conceptual change than a more complex and less appealing model with twelve variables. This final model suggests that instead of focusing on many student characteristics at once, only a few, specifically, three student characteristics truly matters in bringing about conceptual change.

Implications for Practice

Teachers typically have opinions about the role they believe certain student characteristics play in learning. For instance, teachers largely agree students who are more motivated are more likely to learn. That is, students who genuinely have a personal interest in a topic or who feel confident they can learn the content are more likely to do well. Also, teachers generally agree students who have a knack for learning science are more likely to learn difficult and challenging topics. However, there is probably less consensus among teachers concerning the question, do students readily correct principles and concepts they have learned incorrectly? This study suggests one student characteristic, formal reasoning ability, may provide some insight into how willing students are to change their minds and correct misconceptions. But other questions remain, can teachers improve students' formal reasoning ability through instruction? Will increased formal reasoning skills lead to a detectable positive change in conceptual understanding? Shifting teaching practice to improving students' ability to isolate and control variables, as well as use proportional, probabilistic, correlational, and combinatorial reasoning in all indications from these findings should bring about greater success in correcting misconceptions. We do not yet know if formal reasoning ability can be improved through instruction and if enhanced formal reasoning ability will result in greater conceptual change. Nonetheless, this is a natural next step that we should explore in science instruction. Another important implication of this study for practice is teachers should strive to incite students' interest in the subject matter they teach. Perhaps more critically, instructors teaching a diverse student body, whether largely or only slightly so, should endeavor to set the tone for respect and acceptance of others not like themselves. If they are able, teachers should try to reach out to students who seem left out or disenfranchised to help them become active and engaged participants in the learning process.

Limitations of Study

This study involves a sample of introductory biology students in six separate courses at six different postsecondary institutions, with one instructor at each postsecondary institution, for which they are certain limitations. Students were asked to complete four distinct instruments on three discrete occasions over the duration of the unit of evolution, or in some cases, over the course of the semester. Not every student enrolled in the six courses at the six institutions provided a complete set of data on all the variables investigated. Based on the course grade earned, generally, students with higher grades were more likely to complete all the required instruments and provide a full set of data for analyses. Therefore, one limitation is the results probably reflect a more favorable conceptual change than is expected if the response rate were 100%. The expectation is students with higher grades are probably more likely to correct their misconceptions, although this may not necessarily be the case. Every effort was made to choose instruments that are reliable and valid. While the best available instruments were used for this study, no instrument perfectly measures what it is designed to do. Hence, another limitation is instrument design. The inferences drawn from this study are only as good as the instruments used to measure the variables. This study investigated thirteen fixed effects and one random effect. There is still the question of whether an important variable that is related to conceptual change was missed or not considered. The design of the study presented constraints and limitations. The study required the collection of multiple sets of data in a single course that invariably restricted the measurement of some variables. The study did not account for students' content knowledge in relevant areas such as genetics. Also, the study strategies students use, or the effort students put into learning in the course, or the seriousness with which students completed the required instruments were not measured or controlled for in the study.

Recommendations for Future Research

A deeper analysis of the relationship among formal reasoning ability, self-efficacy, intrinsic motivation, pre-instruction score, and post-instruction score is needed. There seems to be a complicated relationship among these variables that warrants further examination. For instance, is there an indirect effect such that formal reasoning ability predicts self-efficacy, which in turn predicts post-instruction scores with pre-instruction score and formal reasoning ability as covariates? Does self-efficacy predict intrinsic motivation, or vice versa? Which component of motivation directly predicts post-instruction score? Is pre-instruction score a covariate with intrinsic motivation or self-efficacy? One recommended course of action is to carry out structural equation modeling analyses to examine the relationship among these variables.

Another pressing question is, what is the nature of the relationship between formal reasoning ability and conceptual change? More precisely, is there another variable, possibility generic cognitive ability to learn, that explains conceptual change? It is certainly instructive to explore whether controlling for generic cognitive ability to learn changes the relationship between formal reasoning ability and conceptual change. This would provide valuable information about whether formal reasoning ability or a more generic cognitive ability to learn is directly related to conceptual change. This line of inquiry would also provide some insight into the likely effects of instruction, if any, on improving formal reasoning skills. Meaning, is instruction to enhance formal reasoning ability a viable option? But an important question emerges, what instrument would be most suitable to measure generic cognitive ability to learn? One possibility is a general intelligence test. The challenge with this approach is often there is a stigma associated with intelligence tests, plus students often find them difficult. Considerable care would be required in designing this study. It is conceivable that IRB would require some assurances that students would not perceive an intelligence test as a threat, deterrent, or distraction from learning in the course in which the study is conducted.

Finally, this study investigates one big idea in one content area in science. A clear recommendation for future research is to expand the inquiry of the relationship of formal reasoning ability, academic motivation, and conceptual change into other big ideas in biology, as well as to other content areas in science. It is important to ask, for example, are formal reasoning ability and academic motivation related to conceptual change in biological systems? Similarly, are formal reasoning ability and academic motivation related to conceptual change for the enduring understanding of chemical reactions in chemistry, or the laws of conservation in physics? Exploring these questions would necessitate the development of valid and reliable instruments to adequately and accurately measure students' conceptual understanding of these enduring ideas in science.

Concluding Remarks

This study makes two important contributions to the study of conceptual change. Firstly, this study highlights the importance of investigating student characteristics, which encompasses rational (formal reasoning ability) and extrarational (academic motivation) aspects of the process of conceptual change. Prior research on conceptual change has neglected to incorporate formal reasoning ability as an important rational component; that is, a cognitive construct. The final stage of Piaget's theory of cognitive development is formal operational thinking (vis-à-vis formal reasoning ability), which was the primary predictor of conceptual change in this study. Secondly, this study took a multidisciplinary perspective by examining formal reasoning ability from cognitive psychology, academic motivation from educational psychology, and teachers' characteristics from science education. These factors from the different disciplinary domains were shown to be important predictors of students' conceptual change.

REFERENCES

Adams, R. J. (2005). Reliability as a measurement design effect. *Studies in Educational Evaluation, 31*(2), 162-172. doi:https://doi.org/10.1016/j.stueduc.2005.05.008

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*(6), 716-723. doi:10.1109/TAC.1974.1100705

American Association for the Advancement of Science (AAAS). (2011). *Vision and change in undergraduate biology education: A call to action* (978-0-87168-741-8). Retrieved from AAAS: http://visionandchange.org/finalreport/

American Association for the Advancement of Science (AAAS). (2013). *Describing and measuring undergraduate STEM teaching practices* (978-0-87168-756-2). Retrieved from http://ccliconference.org/files/2013/11/Measuring-STEM-Teaching-Practices.pdf

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). Standards for educational and psychological testing. Washington, DC: American Educational Research Association.

Anderson, D. L., Fisher, K. M., & Norman, G. J. (2002). Development and evaluation of the conceptual inventory of natural selection. *Journal of Research in Science Teaching, 39*(10), 952-978. doi:10.1002/tea.10053

Andrews, T. M., Leonard, M. J., Colgrove, C. A., & Kalinowski, S. T. (2011). Active learning not associated with student learning in a random sample of college biology courses. *CBE—Life Sciences Education, 10*(4), 394-405. doi:10.1187/cbe.11-07-0061

Arnold, J. B. (2017). ggthemes: Extra themes, scales and geoms for ‘ggplot2’. R package version 3.4.0. Retrieved from https://CRAN.R-project.org/package=ggthemes

Bandura, A. (1989). Human agency in social cognitive theory. *American Psychologist, 44*(9), 1175-1184. *doi:*http://dx.doi.org/10.1037/0003-066X.44.9.1175

Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational Psychologist, 28*(2), 117-148. doi:10.1207/s15326985ep2802_3

Bandura, A. (1994). Self-efficacy. In V. S. Ramachaudran (Ed.), *Encyclopedia of human behavior* (Vol. 4, pp. 71-81). New York, NY: Academic Press.

Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology, 52*(1), 1-26. doi:10.1146/annurev.psych.52.1.1

Bartoń, K. (2018). MuMIn: Multi-model inference: R package version 1.40.4. Retrieved from https://CRAN.R-project.org/package=MuMIn

Berry, A., Friedrichsen, P., & Loughran, J. (Eds.). (2015). *Re-examining pedagogical content knowledge in science education*. New York, NY: Routledge.

Biggs, J. (1993). What do inventories of students' learning processes really measure? A theoretical review and clarification. *British Journal of Educational Psychology, 63*(1), 3-19. doi:10.1111/j.2044-8279.1993.tb01038.x

Biggs, J. B. (1989). Approaches to the enhancement of tertiary teaching. *Higher Education Research & Development*, *8*(1), 7-25. doi:10.1080/0729436890080102

Bishop, B. A., & Anderson, C. W. (1990). Student conceptions of natural selection and its role in evolution. *Journal of Research in Science Teaching, 27*(5), 415-427.

Bonate, P. L. (2000). *Analysis of pretest-posttest designs*. Boca Raton, FL: Chapman & Hall/CRC.

Bruner, J. S. (1959). Inhelder and Piaget's *The growth of logical thinking*. *British Journal of Psychology, 50*(4), 363-370. doi:10.1111/j.2044-8295.1959.tb00716.x

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research, 33*(2), 261-304. doi:10.1177/0049124104268644

Burnham, K. P., Anderson, D. R., & Huyvaert, K. P. (2011). AIC model selection and multimodel inference in behavioral ecology: Some background, observations, and comparisons. *Behavioral Ecology and Sociobiology, 65*(1), 23-35. doi:http://dx.doi.org/10.1007/s00265-010-1029-6

Carey, S. (1992). The origin and evolution of everyday concepts. In R. N. Giere (Ed.), *Minnesota studies in the philosophy of science: Vol. XV. Cognitive models of science* (pp. 89-128). Minneapolis, MN: University of Minnesota Press.

Carey, S., & Spelke, E. (1994). Domain-specific knowledge and conceptual change. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 169-200). Cambridge, UK: Cambridge University Press.

Case, R. (1987). Neo-Piagetian theory: Retrospect and prospect. *International Journal of Psychology, 22*(5/6), 773-791. doi:10.1080/00207598708246802

Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, 48(6), 1-29. doi:10.18637/jss.v048.i06

Chalmers, R. P. (2015a, Oct 12, 2015). Is empirical reliability in IRT the counterpart of internal consistency in CTT? Retrieved from https://stats.stackexchange.com/questions/173988/is-empirical-reliability-in-irt-the-counterpart-of-internal-consistency-in-ctt

Chalmers, R. P. (2015b, Jan 10, 2015). Reliability in IRT style. Retrieved from https://stats.stackexchange.com/questions/132738/reliability-in-irt-style

Chalmers, R. P. (2018, Aug 10, 2018). R Package Documentation: philchalmers/mirt documentation. Retrieved from https://rdrr.io/github/philchalmers/mirt/man/empirical_rxx.html

Chen, X. (2013). STEM Attrition: *College students' paths into and out of STEM fields (NCES 2014-001)*. National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. Retrieved from https://files.eric.ed.gov/fulltext/ED544470.pdf

Chi, M. T. H. (1992). Conceptual change within and across ontological categories: Examples from learning and discovery in science. In R. N. Giere (Ed.), *Minnesota studies in the philosophy of science: Vol. XV. Cognitive models of science* (pp. 129-186). Minneapolis, MN: University of Minnesota Press.

Chi, M. T. H. (2008). Three types of conceptual change: Belief revision, mental model transformation, and categorical shift. In S. Vosniadou (Ed.), *International handbook of research on conceptual change* (pp. 61-82). New York, NY: Routledge.

Chism, N. (1994). Taking student diversity into account. In W. J. McKeachie (Ed.), *Teaching tips: Strategies, research, and theory for college university teachers* (9th ed., pp. 223-237). Lexington, MA: D.C. Heath & Co.

Clement, J. (2008). The role of explanatory models in teaching for conceptual change. In S. Vosniadou (Ed.), *International handbook of research on conceptual change* (pp. 417-452). New York, NY: Routledge.

Collins, J. B., & Pratt, D. D. (2010). The teaching perspectives inventory at 10 Years and 100,000 respondents: Reliability and validity of a teacher self-report inventory. *Adult Education Quarterly*, *61*(4), 358-375. doi:10.1177/0741713610392763

Confrey, J. (1990). A review of the research on student conceptions in mathematics, science, and programming. *Review of Research in Education, 16*(1), 3-56.

Cronbach, L. J., & Furby, L. (1970). How we should measure "change": Or should we? *Psychological Bulletin, 74*(1), 68-80. doi:http://dx.doi.org/10.1037/h0029382

Dancy, M., & Henderson, C. (2007). Framework for articulating instructional practices and conceptions. *Physical Review Special Topics - Physics Education Research, 3*(1), 010103. doi:10.1103/PhysRevSTPER.3.010103

Dictionary.com, LLC. (2018). Definitions. Retrieved from http://www.dictionary.com/browse/scientific-theory

Dimitrov, D. M., & Rumrill, P. D. (2003). Pretest-posttest designs and measurement of change. *Work (Reading, Mass.), 20*(2), 159-165.

diSessa, A. A. (2006). A history of conceptual change research: Threads and fault lines. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 265-281). Cambridge, UK: Cambridge University Press.

Dobzhansky, T. (1973). Nothing in biology makes sense except in the light of evolution. *The American Biology Teacher, 35*(3), 125-129.

Dole, J. A., & Sinatra, G. M. (1998). Reconceptalizing change in the cognitive construction of knowledge. *Educational Psychologist, 33*(2-3), 109-128. doi:10.1080/00461520.1998.9653294

Doran, R. L. (1972). Misconceptions of selected science concepts held by elementary school students. *Journal of Research in Science Teaching, 9*(2), 127-137.

Duit, R., Treagust, D. F., & Widodo, A. (2008). Teaching science for conceptual change: Theory and practice. In S. Vosniadou (Ed.), *International handbook of research on conceptual change* (pp. 629-646). New York, NY: Routledge.

Farber, P. (2003). Teaching evolution and the nature of science. *The American Biology Teacher, 65*(5), 347-354. doi:10.1662/0002-7685(2003)065[0347:tetnos]2.0.co;2

Fennema, E., & Sherman, J. (1977). Sex-related differences in mathematics achievement, spatial visualization and affective factors. *American Educational Research Journal, 14*(1), 51-71. doi:10.3102/00028312014001051

Flavell, J. H. (1992). Cognitive development: Past, present, and future. *Developmental Psychology, 28*(6), 998-1005. doi:http://dx.doi.org/10.1037/0012-1649.28.6.998

Fox, J. (2003). Effect displays in R for generalised linear models. *Journal of Statistical Software*, 8(15), 1-27. Retrieved from http://www.jstatsoft.org/v08/i15/

Fox, J., & Weisberg, S. (2011). An {R} Companion to Applied Regression. Thousand Oaks, CA: Sage. Retrieved from http://socserv.socsci.mcmaster.ca/jfox/Books/Companion

Galecki, A., & Burzykowski, T. (2013). *Linear mixed-effects models using R: A step-by-step approach*. New York, NY: Springer.

Glynn, S. M., Brickman, P., Armstrong, N., & Taasoobshirazi, G. (2011). Science motivation questionnaire II: Validation with science majors and nonscience majors. *Journal of Research in Science Teaching, 48*(10), 1159-1176. doi:10.1002/tea.20442

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

Glynn, S. M., Taasoobshirazi, G., & Brickman, P. (2007). Nonscience majors learning science: A theoretical model of motivation. *Journal of Research in Science Teaching, 44*(8), 1088-1107. doi:10.1002/tea.20181

Glynn, S. M., Taasoobshirazi, G., & Brickman, P. (2009). Science motivation questionnaire: Construct validation with nonscience majors. *Journal of Research in Science Teaching, 46*(2), 127-146. doi:10.1002/tea.20267

Greenwood, M. C. (2017). *STAT 448 fall class materials with function to calculate variance inflation factor (VIF) for mixed models*. Mathematical Sciences. Montana State University. Bozeman, MT.

Gregory, T. R. (2009). Understanding natural selection: Essential concepts and common misconceptions. *Evolution: Education and Outreach, 2*(2), 156-175. doi:10.1007/s12052-009-0128-1

Halpern, D. F. (2004). A cognitive-process taxonomy for sex differences in cognitive abilities. *Current Directions in Psychological Science, 13*(4), 135-139. doi:10.1111/j.0963-7214.2004.00292.x

Henderson, C., & Dancy, M. H. (2007). Barriers to the use of research-based instructional strategies: The influence of both individual and situational characteristics. *Physical Review Special Topics - Physics Education Research, 3*(2), 020102.

Higher Education Research Institute (HERI). (2010). *Degree of success: Bachelor's degree completion rates among initial STEM majors*. Retrieved from University of California Los Angeles (UCLA)*:* https://heri.ucla.edu/nih/downloads/2010-Degrees-of-Success.pdf

Hynd, C. (2003). Conceptual change in response to persuasive messages. In G. M. Sinatra & P. R. Pintrich (Eds.), *Intentional conceptual change* (pp. 291-315). Mahwah, NJ: Lawrence Erlbaum.

Indiana University Center for Postsecondary Research (IUCPR). (2017). The Carnegie Classification of Institutions of Higher Education (CCIHE)®. Retrieved from http://carnegieclassifications.iu.edu

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning with applications in R*. In G. Casella, S. E. Fienberg, & I. Olkin (Series Eds.), Springer texts in statistics. doi:10.1007/978-1-4614-7138-7

Johnson, M. A., & Lawson, A. E. (1998). What are the relative effects of reasoning ability and prior knowledge on biology achievement in expository and inquiry classes? *Journal of Research in Science Teaching, 35*(1), 89-103.

Jonassen, D. (2008). Model building for conceptual change. In S. Vosniadou (Ed.), *International handbook of research on conceptual change* (pp. 676-693). New York, NY: Routledge.

Kalinowski, S. T., Leonard, M. J., Andrews, T. M., & Litt, A. R. (2013). Six classroom exercises to teach natural selection to undergraduate biology students. *CBE—Life Sciences Education, 12*(3), 483–493. doi:10.1091/cbe-12-06-0070

Kalinowski, S. T., Leonard, M. J., & Taper, M. L. (2016). Development and validation of the conceptual assessment of natural selection (CANS). *CBE—Life Sciences Education, 15*(4). doi:10.1187/cbe.15-06-0134

Kalinowski, S. T., Taper, M. L., Leonard, M. J., & Willoughby, S. D. (2018). Development and validation of a scientific reasoning test for college students. *Submitted*.

Karunanayake, D., & Nauta, M. M. (2004, 2004/03//). The relationship between race and students' identified career role models and perceived role model influence. *Career Development Quarterly, 52*, 225-234.

Kober, N. (2015). *Reaching students: What research says about effective instruction in undergraduate science and engineering*. Retrieved from https://www.nap.edu/download/18687 doi:10.17226/18687

Kokkelenberg, E. C., & Sinha, E. (2010). Who succeeds in STEM studies? An analysis of Binghamton University undergraduate students. *Economics of Education Review, 29*(6), 935-946. doi:https://doi.org/10.1016/j.econedurev.2010.06.016

Kuhn, T. S. (1970). *Structure of scientific revolutions* (2nd ed. Vol. 2). Chicago, IL: The University of Chicago Press.

Lawson, A. E. (1978). The development and validation of a classroom test of formal reasoning. *Journal of Research in Science Teaching, 15*(1), 11-24. doi:10.1002/tea.3660150103

Lawson, A. E. (1985). A review of research on formal reasoning and science teaching. *Journal of Research in Science Teaching, 22*(7), 569-617.

Lawson, A. E. (1992). What do tests of "formal" reasoning actually measure? *Journal of Research in Science Teaching, 29*(9), 965-983. doi:10.1002/tea.3660290906

Lawson, A. E. (2000a). *Classroom test of scientific reason: Revised pencil-paper edition*. Arizona State University. Tempe, AZ.

Lawson, A. E. (2000b). The generality of hypothetico-deductive reasoning: Making scientific thinking explicit. *The American Biology Teacher, 62*(7), 482-495.

Lawson, A. E. (2004). The nature and development of scientific reasoning: A synthetic view. *International Journal of Science and Mathematics Education, 2*(3), 307–338.

Lawson, A. E., Adi, H., & Karplus, R. (1979). Development of correlational reasoning in secondary schools: Do biology courses make a difference? *The American Biology Teacher, 41*(7), 420-430. doi:10.2307/4446678

Lawson, A. E., Banks, D. L., & Logvin, M. (2007). Self-efficacy, reasoning ability, and achievement in college biology. *Journal of Research in Science Teaching, 44*(5), 706-724. doi:10.1002/tea.20172

Lawson, A. E., Clark, B., Cramer-Meldrum, E., Falconer, K. A., Sequist, J. M., & Kwon, Y. J. (2000). Development of scientific reasoning in college biology: Do two levels of general hypothesis-testing skills exist? *Journal of Research in Science Teaching, 37*(1), 81-101. doi:10.1002/(SICI)1098-2736(200001)37:1<81::AID-TEA6>3.0.CO;2-I

Lawson, A. E., & Renner, J. W. (1975). Relationships of science subject matter and developmental levels of learners. *Journal of Research in Science Teaching, 12*(4), 347-358. doi:10.1002/tea.3660120405

Lawson, A. E., & Thompson, L. D. (1988). Formal reasoning ability and misconceptions concerning genetics and natural selection. *Journal of Research in Science Teaching, 25*(9), 733-746. doi:10.1002/tea.3660250904

Leonard, M. J., Kalinowski, S. T., & Andrews, T. M. (2014). Misconceptions yesterday, today, and tomorrow. *CBE—Life Sciences Education, 13*(2), 179-186. doi:10.1187/cbe.13-12-0244

Limón, M. (2001). On the cognitive conflict as an instructional strategy for conceptual change: A critical appraisal. *Learning and Instruction, 11*(4), 357-380. doi:https://doi.org/10.1016/S0959-4752(00)00037-2

Linn, M. C. (2008). Teaching for conceptual change: Distinguish or extinguish ideas. In S. Vosniadou (Ed.), *International handbook of research on conceptual change* (pp. 695-722). New York, NY: Routledge.

Linnenbrink-Garcia, L., Pugh, K. J., Koskey, K. L. K., & Stewart, V. C. (2012). Developing conceptual understanding of natural selection: The role of interest, efficacy, and basic prior knowledge. *The Journal of Experimental Education, 80*(1), 45-68. doi:10.1080/00220973.2011.559491

Linnenbrink, E. A., & Pintrich, P. R. (2003). Achievement goals and intentional conceptual change. In G. M. Sinatra & P. R. Pintrich (Eds.), *Intentional conceptual change* (pp. 343-369). Mahwah, NJ: Lawrence Erlbaum.

Loughran, J., Mulhall, P., & Berry, A. (2004). In search of pedagogical content knowledge in science: Developing ways of articulating and documenting professional practice. *Journal of Research in Science Teaching, 41*(4), 370-391.

McKeachie, W. J., Pintrich, P. R., Lin, Y.-G., & Smith, D. A. F. (1986). *Teaching and learning in the college classroom: A review of the research literature*. Ann Arbor, MI: National Center for Research to Improve Postsecondary Teaching and Learning.

Midgley, C., Maehr, M. L., Hruda, L. Z., Anderman, E., Anderman, L., Freeman, K. E., . . . Urdan, T. (2000). *Manual for the Patterns of Adaptive Learning Scales (PALS).* Retrieved from Ann Arbor, MI: The University of Michigan

Mills, G. E., & Gay, L. R. (2016). *Educational research: Competencies for analysis and applications* (11th, Loose-Leaf Version ed.). Boston, MA: Pearson.

Mintzes, J. J., Trowbridge, J. E., Arnaudin, M. W., & Wandersee, J. H. (1991). Children's biology: Studies in conceptual development in the life sciences. In S. M. Glynn, R. H. Yeany, & B. K. Britton (Eds.), *The psychology of learning science* (pp. 190-214). New York, NY: Routledge.

Murphy, P. K., & Alexander, P. A. (2008). The role of knowledge, beliefs, and interest in the conceptual change process: A synthesis and meta-analysis of the research. In S. Vosniadou (Ed.), *International handbook of research on conceptual chang* (pp. 583-616). New York, NY: Routledge.

National Center for Education Statistics (NCES), (2016, May). The condition of education: High school coursetaking. Retrieved from https://nces.ed.gov/programs/coe/indicator_cod.asp

National Center for Education Statistics (NCES). (2018). Integrated Postsecondary Education Data System (IPEDS). Retrieved from https://nces.ed.gov/ipeds/

Nehm, R. H., Beggrow, E. P., Opfer, J. E., & Ha, M. (2012). Reasoning about natural selection: Diagnosing contextual competency using the ACORNS instrument. *The American Biology Teacher, 74*, 92+.

Nehm, R. H., & Reilly, L. (2007). Biology majors' knowledge and misconceptions of natural selection. *BioScience, 57*(3), 263-272. doi:10.1641/B570311

Nehm, R. H., & Schonfeld, I. S. (2008). Measuring knowledge of natural selection: A comparison of the CINS, an open-response instrument, and an oral interview. *Journal of Research in Science Teaching, 45*(10), 1131-1160. doi:10.1002/tea.20251

Nelson, D. J., & Brammer, C. N. (2010). *A national analysis of minorities in science and engineering faculties at research universities*. Retrieved from http://faculty-staff.ou.edu/N/Donna.J.Nelson-1/diversity/Faculty_Tables_FY07/07Report.pdf

Patrick, H., Ryan, A. M., & Pintrich, P. R. (1999). The differential impact of extrinsic and mastery goal orientations on males' and females' self-regulated learning. *Learning and Individual Differences*, *11*(2), 153-171.

Perna, L., Lundy-Wagner, V., Drezner, N. D., Gasman, M., Yoon, S., Bose, E., & Gary, S. (2009). The contribution of HBCUS to the preparation of African American women for STEM careers: A case study. *Research in Higher Education, 50*(1), 1-23. doi:10.1007/s11162-008-9110-y

Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 19, pp. 123-205). Cambridge, MA: Academic Press.

Piaget, J. (1952). *The origins of intelligence in children*. New York, NY: International Universities Press.

Piaget, J. (1964). Part I: Cognitive development in children: Piaget development and learning. *Journal of Research in Science Teaching, 2*(3), 176-186. doi:10.1002/tea.3660020306

Piaget, J. (1972). Intellectual evolution from adolescence to adulthood. *Human Development*, *15*(1), 1-12.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team. (2018). nlme: Linear and nonlinear mixed effects models: R package version 3.1-131.1. Retrieved from https://CRAN.R-project.org/package=nlme

Pintrich, P. R. (1999). Motivational beliefs as resources for and constraints on conceptual change. In W. Schnotz, S. Vosniadou, & M. Carretero (Eds.), *New perspectives on conceptual change* (pp. 33-50). Bingley, UK: Emerald Group Publishing.

Pintrich, P. R., Marx, R. W., & Boyle, R. A. (1993). Beyond cold conceptual change: The role of motivational beliefs and classroom contextual factors in the process of conceptual change. *Review of Educational Research*, *63*(2), 167-199.

Pintrich, P. R., & Sinatra, G. M. (2003). Future directions for theory and research on intentional conceptual change. In G. M. Sinatra & P. R. Pintrich (Eds.), *Intentional conceptual change* (pp. 423-434). Mahwah, NJ: Lawrence Erlbaum

Pintrich, P. R., Smith, D. A. F., Garcia, T., & McKeachie, W. J. (1991). *A Manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ)* (NCRIPTAL-91-B-004). Retrieved from Ann Arbor, MI: The University of Michigan

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

Price, J. (2010). The effect of instructor race and gender on student persistence in STEM fields. *Economics of Education Review, 29*(6), 901-910. doi:https://doi.org/10.1016/j.econedurev.2010.07.009

Prosser, M., & Trigwell, K. (2006). Confirmatory factor analysis of the Approaches to Teaching Inventory. *British Journal of Educational Psychology, 76*(2), 405-419. doi:10.1348/000709905X43571

Prosser, M., Trigwell, K., & Taylor, P. (1994). A phenomenographic study of academics' conceptions of science teaching and learning. *Learning and Instruction, 4*(3), 217-231.

Pruim, R., Kaplan, D. T., & Horton, N. J. (2017). The mosaic package: Helping students to 'think with data' using R. *The R Journal*, 9(1), 77-102. Retrieved from https://journal.r-project.org/archive/2017/RJ-2017-024/index.html

R Core Team. (2016). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Ramsden, P. (1991). A performance indicator of teaching quality in higher education: The Course Experience Questionnaire. *Studies in Higher Education, 16*(2), 129-150. doi:10.1080/03075079112331382944

Ramsey, F. L., & Schafer, D. W. (2013). *The statistical sleuth: A course in methods of data analysis*. Boston, MA: Brooks/Cole Cengage Learning.

Rask, K. (2010). Attrition in STEM fields at a liberal arts college: The importance of grades and pre-collegiate preferences. *Economics of Education Review, 29*(6), 892-900. doi:https://doi.org/10.1016/j.econedurev.2010.06.013

Reeve, J., Nix, G., & Hamm, D. (2003). Testing models of the experience of self-determination in intrinsic motivation and the conundrum of choice. *Journal of Educational Psychology, 95*(2), 375-392. doi:http://dx.doi.org/10.1037/0022-0663.95.2.375

Rendon, L. I. (1994). Validating culturally diverse students: Toward a new model of learning and student development. *Innovative Higher Education, 19*(1), 33-51. doi:10.1007/bf01191156

Revelle, W. (2017). psych: Procedures for Personality and Psychological Research. Evanston, IL: Northwestern University. Retrieved from https://CRAN.R-project.org/package=psych Version = 1.7.8.

Roberts, R. S. (1980). Concurrent validity in tests of piagetian developmental levels. *Journal of Research in Science Teaching, 17*(4), 343-350. doi:10.1002/tea.3660170412

Rosenblatt, L. (Ed.) (2011). *Rethinking the way we teach science: The interplay of content, pedagogy, and the nature of science*. New York, NY: Routledge.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1-36. Retrieved from URL http://www.jstatsoft.org/v48/i02/

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, *25*(1), 54-67. doi:10.1006/ceps.1999.1020

Sayre, S., & Ball, D. W. (1975). Piagetian cognitive development and achievement in science. *Journal of Research in Science Teaching, 12*(2), 165-174.

Schunk, D. H. (2016). *Learning theories: An educational perspective* (Loose-leaf version, 7th ed. ed.). Boston, MA: Pearson.

Schunk, D. H., Meece, J. L., & Pintrich, P. R. (2014). *Motivation in education: Theory, research, and applications* (4th ed.). Boston, MA: Pearson.

Shayer, M., Adey, P., & Wylam, H. (1981). Group tests of cognitive development ideals and a realization. *Journal of Research in Science Teaching, 18*(2), 157-168. doi:10.1002/tea.3660180208

Siegel, R. S., & Svetina, M. (2008). Relationships between short- and long-term changes in children's thinking. In S. Vosniadou (Ed.), *International handbook of research on conceptual change* (pp. 102-123). New York, NY: Routledge.

Sinatra, G. M. (2005). The “warming trend” in conceptual change research: The legacy of Paul R. Pintrich. *Educational Psychologist, 40*(2), 107-115. doi:10.1207/s15326985ep4002_5

Sinatra, G. M., Brem, S. K., & Evans, E. M. (2008). Changing minds? Implications of conceptual change for teaching and learning about biological evolution. *Evolution: Education and Outreach, 1*(2), 189-195. doi:10.1007/s12052-008-0037-8

Sinatra, G. M., & Mason, L. (2008). Beyond knowledge: Learner characteristics influencing conceptual change. In S. Vosniadou (Ed.), *International handbook of research on conceptual change* (pp. 560-582). New York, NY: Routledge.

Sinatra, G. M., & Pintrich, P. R. (2003a). The role of intentions in conceptual change learning. In G. M. Sinatra & P. R. Pintrich (Eds.), Intentional conceptual change (pp. 1-17). Mahwah, NJ: Lawrence Erlbaum.

Sinatra, G. M., & Pintrich, P. R. (Eds.). (2003b). Intentional conceptual change. Mahwah, NJ: Lawrence Erlbaum.

Sinatra, G. M., Southerland, S. A., McConaughy, F., & Demastes, J. W. (2003). Intentions and beliefs in students' understanding and acceptance of biological evolution. *Journal of Research in Science Teaching, 40*(5), 510-528. doi:10.1002/tea.10087

Singer, S. R., Nielsen, N. R., & Schweingruber, H. A. (2012). *Discipline-based education research: Understanding and improving learning in undergraduate science and engineering* (978-0-309-25411-3). Retrieved from National Academy of Sciences: http://nap.edu/13362

Solorzano, D., Ceja, M., & Yosso, T. (2000). Critical Race Theory, racial microaggressions, and campus racial climate: The experiences of African American college students. *The Journal of Negro Education, 69*(1/2), 60-73.

Southerland, S. A., & Sinatra, G. M. (2003). Learning about biological evolution: A special case of intentional conceptual change. In G. M. Sinatra & P. R. Pintrich (Eds.), *Intentional conceptual change* (pp. 315-342). Mahwah, NJ: Lawrence Erlbaum.

Tobin, K. G., & Capie, W. (1981). The development and validation of a group test of logical thinking. *Educational and Psychological Measurement*, *41*(2), 413-423. doi:10.1177/001316448104100220

Tobin, K. G., & Capie, W. (1982). Development and validation of a group test of integrated science processes. *Journal of Research in Science Teaching, 19*(2), 133-141. doi:10.1002/tea.3660190205

Touré-Tillery, M., & Fishbach, A. (2014). How to Measure Motivation: A Guide for the Experimental Social Psychologist. *Social and Personality Psychology Compass, 8*(7), 328-341. doi:10.1111/spc3.12110

Treisman, U. (1992). Studying students studying calculus: A look at the lives of minority mathematics students in college. *The College Mathematics Journal*, *23*(5), 362-372. doi:10.2307/2686410

Trigwell, K., & Prosser, M. (1996a). Changing approaches to teaching: A relational perspective. *Studies in Higher Education, 21*(3), 275-284. doi:10.1080/03075079612331381211

Trigwell, K., & Prosser, M. (1996b). Congruence between intention and strategy in university science teachers' approaches to teaching. *Higher Education, 32*(1), 77-87. doi:10.1007/bf00139219

Trigwell, K., & Prosser, M. (2004). Development and use of the approaches to teaching inventory. *Educational Psychology Review, 16*(4), 409-424. doi:10.1007/s10648-004-0007-9

Trigwell, K., Prosser, M., & Taylor, P. (1994). Qualitative differences in approaches to teaching first year university science. *Higher Education, 27*(1), 75-84. doi:10.1007/bf01383761

Trigwell, K., Prosser, M., & Waterhouse, F. (1999). Relations between teachers' approaches to teaching and students' approaches to learning. *Higher Education, 37*(1), 57-70. doi:10.1023/a:1003548313194

Urry, L. A., Cain, M. A., Wasserman, S. A., Minorsky, P. V., & Reece, J. B. (2017). *Campbell biology* (11th ed.). New york, NY: Pearson.

Valentine, J. C., DuBois, D. L., & Cooper, H. (2004). The relation between self-beliefs and academic achievement: A meta-analytic review. *Educational Psychologist, 39*(2), 111-133. doi:10.1207/s15326985ep3902_3

Vallerand, R. J., Pelletier, L. G., Blais, M. R., Briere, N. M., Senecal, C., & Vallieres, E. F. (1992). The academic motivation scale: A measure of intrinsic, extrinsic, and amotivation in education. *Educational and Psychological Measurement, 52*(4), 1003-1017. doi:10.1177/0013164492052004025

Vosniadou, S. (2003). Exploring the relationships between conceptual change and intentional learning. In G. M. Sinatra & P. R. Pintrich (Eds.), *Intentional conceptual change* (pp. 373-402). Mahwah, NJ: Lawrence Erlbaum.

Vosniadou, S. (2008a). Conceptual change research: An introduction. In S. Vosniadou (Ed.), *International handbook of research on conceptual change* (pp. xiii-xxviii). New York, NY: Routledge.

Vosniadou, S. (2014). Examining cognitive development from a conceptual change point of view: The framework theory approach. *European Journal of Developmental Psychology, 11*(6), 645-661. doi:10.1080/17405629.2014.921153

Vosniadou, S. (Ed.) (2008b). *International handbook of research on conceptual change*. New York, NY: Routledge.

Vosniadou, S. (Ed.) (2013). *International handbook of research on conceptual change* (2nd ed.). New York, NY: Routledge.

Vosniadou, S., & Ioannides, C. (1998). From conceptual development to science education: a psychological point of view. *International Journal of Science Education, 20*(10), 1213-1230. doi:10.1080/0950069980201004

Weiner, B. (1972). Attribution theory, achievement motivation, and the educational process. *Review of Educational Research, 42*(2), 203-215. doi:10.3102/00346543042002203

Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological Review, 92*(4), 548-573. doi:http://dx.doi.org/10.1037/0033-295X.92.4.548

White, R. T., & Gunstone, R. F. (2008). The conceptual change approach and the teaching of science. In S. Vosniadou (Ed.), *International handbook of research on conceptual change* (pp. 619-628). New York, NY: Routledge.

Wickham, H. (2007). Reshaping Data with the reshape Package. *Journal of Statistical Software*, 21(12), 1-20. Retrieved from http://www.jstatsoft.org/v21/i12/

Wickham, H. (2009). ggplot2: Elegant graphics for data analysis. New York, NY: Springer-Verlag. Retrieved from http://ggplot2.org

Wickham, H., & Bryan, J. (2017). readxl: Read excel. R package version 1.0.0. Retrieved from https://CRAN.R-project.org/package=readxl

Wieman, C., & Gilbert, S. (2014). The teaching practices inventory: A new tool for characterizing college and university teaching in mathematics and science. *CBE—Life Sciences Education, 13*(3), 552-569. doi:10.1187/cbe.14-02-0023

Williams, R. H., & Zimmerman, D. W. (1996). Are simple gain scores obsolete? *Applied Psychological Measurement, 20*(1), 59-69. doi:10.1177/014662169602000106

Yosso, T. J., Smith, W. A., Ceja, M., & Solórzano, D. G. (2009). Critical Race Theory, racial microaggressions, and campus racial climate for Latina/o undergraduates. *Harvard Educational Review, 79*(4), 659-690,781,785-786.

Zuur, A. F., Ieno, E. N., Wlaker, N. J., Saveliev, A. A., & Smith, G. M. (2009). *Mixed effects models and extentions in ecology with R*. New York, NY: Springer.

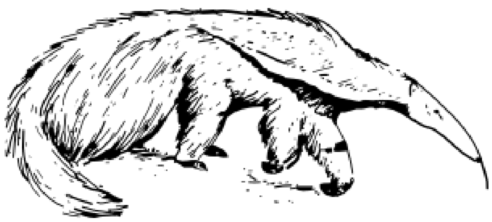
APPENDICES

APPENDIX A

CONCEPTUAL ASSESSMENT OF NATURAL SELECTION (CANS) TEST

Conceptual Assessment of Natural Selection (CANS) Test

The following questions relate to evolution. Please answer the questions as you think an evolutionary biologist would.



Anteaters are mammals that live in South America and eat only ants. Anteaters have several traits that help them catch and eat ants efficiently. Firstly, anteaters have remarkably large claws that allow them to easily rip open ant hills. Anteaters feed by sticking their tongue into tunnels in the ant hills. Their entire head and mouth is adapted for catching ants. Their tongues are 24 inches long and covered with sticky saliva. Anteaters cannot open their mouth, and do not have teeth. Even the stomachs of anteaters are unique: unlike most mammals, anteaters do not secrete acid in their stomachs. None is needed. Ants naturally contain formic acid; ants eaten by anteaters' digest in their own acid. Biologists have concluded that anteaters evolved all of these unique traits from ancestors that looked similar to rats.

1. Which of the following is the best description of how anteaters evolved long tongues?

a. Individuals adapted because they needed to reach ants deep inside anthills.
b. Individuals stretched their tongues to reach ants deep inside anthills.
c. Individuals with the shortest tongues could not catch ants and starved.
d. Random mutations occurred because anteaters needed longer tongues.
e. Changes like this depend on many factors, so it is impossible to answer.

2. Anteaters evolved long claws from ancestors that had shorter claws. Think about the first anteater to have claws as long as modern anteaters. Why did this individual have such long claws?

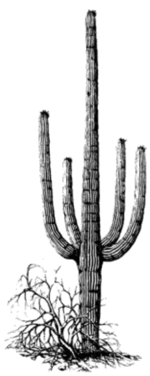
a. The anteater dug up many anthills, and these efforts affected its claws.
b. The anteater was lucky a genetic accident gave it long claws.
c. The anteater needed long claws to dig up ants, so they developed.
d. The anteater needed long claws to eat, so a mutation changed its DNA.

3. An ancestor of modern anteaters has a tongue that is only half as long as modern anteaters. Because her tongue is relative short, she has to work hard to extend her tongue far enough inside ant hills to catch ants. How will these efforts affect the tongues of her offspring?

a. Her efforts will probably give her offspring slightly longer tongues.
b. Her efforts will not affect the tongues of her offspring.
c. Growth is affected by many factors; the effects of her actions cannot be predicted.

4. Modern anteaters do not have teeth, but their ancestors did. Which of the following is the best description of what caused anteaters to lose their teeth?

a. Anteaters did not use their teeth while feeding on ants.
b. Anteaters did not need their teeth to survive and raise their young.
c. Anteaters without teeth had more young than anteaters with teeth.
d. This happened entirely by chance.
e. Changes like this depend on many factors, so it is impossible to answer.



Saguaro cacti live in the scorching hot desserts of Arizona and Mexico where it is common for less than eight inches of rain to fall in a year. Not surprisingly, saguaro cacti have special traits that help them live in this harsh landscape. Like many cacti, saguaro have no leaves. The leaves of the ancestors of saguaro evolved into spines. This prevents water from evaporating from leaves and deters animals from feeding on the saguaro. Photosynthesis occurs within the stem and branches of saguaros. Water loss is further minimized by a waxy covering on the "skin" of saguaro. Beneath the ground, saguaros have an extensive root system—much longer than plants living in wetter climates. These roots are also remarkably shallow. Most of them are only a few inches below the surface of the ground. This enables saguaros to absorb water from rainfall before it evaporates back into the air.

5. What traits do saguaro cacti inherit from their parents?

a. Traits that helped their parents survive and reproduce.
b. Traits that were changed by the environment during their parents' lifetime.
c. Traits that were determined by genes.
d. Traits determined by genes plus one or more other traits listed above.

6. Which of the following is the best description of how saguaro cacti evolved to have long roots?

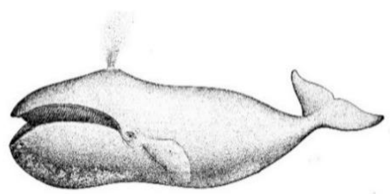
a. Saguaro cacti grew longer roots because they needed to collect water.
b. Mutations occurred because the climate of Arizona and Mexico was hot.
c. Saguaros with short roots did not produce as many seeds as saguaros with longer roots.
d. Changes like this depend on many factors, so it is impossible to answer.

7. Every individual plant and animal is affected by the environment during its lifetime. For example, a person will become tan if exposed to the sun, and a tree will grow slanted if it lives on a windy ridge. What role did the responses of individuals to their environment (like these) play in the evolution of waxy skin among saguaro cacti?

a. Responses like these were the sole reason saguaro cacti evolved waxy skin.
b. Responses like these contributed to saguaro cacti evolving waxy skin.
c. Responses like these played no role in the evolution of waxy skin in saguaro cacti.
d. Responses like these might have played a role in the evolution of cacti—if saguaro respond to intense sunlight or drought by growing waxier skin.

8. The ancestors of modern saguaro cacti did not have long and sharp spines. Consider the first ancestor of saguaro cacti to grow spines that were as long and sharp as the spines on saguaro cacti living in Arizona today. Why did this cactus grow such sharp spines?

a. It was fortunate a genetic mistake gave it extra sharp spines.
b. The cactus needed sharper spines to stop animals from eating it.
c. Animals chewing on the cactus caused it to grow sharper spines.
d. Mutations changed the DNA of this cacti because it was injured by an animal.
e. The hot climate caused this change.



Bowhead whales are the only species of large whales that live their entire life in the icy water of the Arctic Ocean. They have a couple adaptations that help them do this. First, bowhead whales have a thick layer of fat under their skin called blubber that helps keep them warm. The blubber of bowhead whales is 18 inches thick, which is thicker than any other whale. Second, bowhead whales have a remarkably thick skull. This allows them to break thick ice in order to get air to breathe. No other whales have such thick skulls. Available evidence shows that bowhead whales evolved from ancestors that lived in the warm waters of the Pacific Ocean and did not have either thick blubber or thick skulls.

9. Which of the following is the best description of the role cold water played in the evolution of thick blubber?

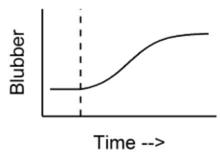
a. It caused mutations that gave whales thicker and thicker blubber.
b. It helped determine which individuals reproduced and which did not.
c. It altered the growth and development of individuals.
d. It forced individuals to adapt in order to survive.

10. Consider a baby whale born during the time the ancestors of modern bowhead whales were evolving thicker skulls. When this whale grows up, how will its skull compare to the skulls of its parents?

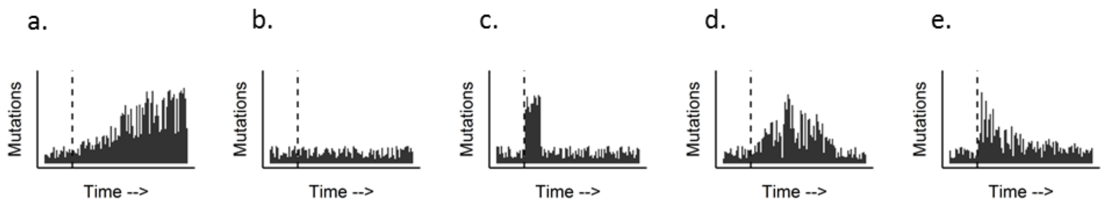
a. Its skull will probably be slightly thicker than the skulls of its parents.
b. Its skull will probably be similar to its parents.
c. Its skull will probably be slightly thinner than the skulls of its parents.
d. Skull development is affected by too many factors to make a prediction

11. What is the best way to describe the evolutionary changes that occurred among the whales while the species evolved thick skulls?

a. The skull of each whale got a little thicker during its lifetime.
b. Whales with thick skulls reproduced and became more common.
c. The population changed randomly each year.
d. Mutations increased the skull thickness of more and more whales each year.



12. The graph at left shows the average blubber thickness of bowhead whales before and after they moved to the Arctic Ocean. Which graph below shows the number of mutations occurring in the population during this time period? (The dashed line shows when the whales moved to the Arctic).



Malaria is a deadly tropical disease caused by a single-celled parasite. People become infected with malaria when they are bitten by a mosquito that carries the parasite. A common approach for preventing malaria is killing the mosquitoes that spread the parasite. In the 1940s the insecticide DDT was discovered to be highly effective at killing mosquitoes. DDT was sprayed in many tropical countries and initially killed 99% percent of the mosquitoes in the areas where it was used.

Soon after DDT was first used, health workers in Africa discovered that mosquito populations evolved to be resistance to DDT: each year DDT was applied, fewer and fewer of the mosquitoes exposed to DDT died. By the end of the 1940s DDT was no longer effective in some regions and had to be replaced with other insecticides. This happened everywhere DDT was used. Switching insecticides proved to be only a temporary fix; mosquitoes evolved resistance to each insecticide that has been used to combat malaria.

13. Which of the following is the best explanation of the process that caused mosquito populations in Africa to become resistant to DDT?

a. The immune systems of mosquitoes exposed to DDT developed resistance. These mosquitoes passed some of this resistance to their offspring so that each generation of mosquitoes became more likely to survive exposure to DDT.
b. Some mosquitoes were fortunate to be naturally able to survive exposure to DDT (even though DDT had never been used in their area before), and passed this ability to their offspring.
c. The mosquitoes became resistant, because if they did not, they would die.
d. DDT caused widespread mutations in the DNA of mosquitoes.

14. Consider a female mosquito that was exposed to DDT during the years a population was evolving resistance to DDT. She survives and later lays a cluster of eggs. How will her exposure to DDT likely affect her offspring?

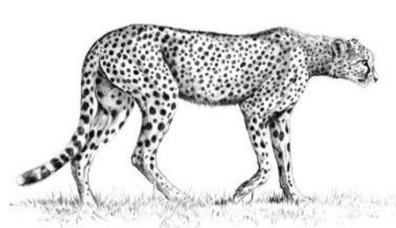
a. Her exposure to DDT will give her offspring increased resistance to DDT.
b. Her exposure to DDT will have no effect on her offspring.
c. The effect of her exposure to DDT on her offspring cannot be predicted.

15. What was most likely true regarding the genetic mutations that occurred during the years mosquitoes were evolving resistance to DDT?

a. The number and effect of mutations that occurred was not influenced by DDT.
b. Most of the mutations that occurred helped the mosquitoes survive.
c. The number of mutations occurring in the population increased when DDT was first applied, and then decreased when the mosquitoes finished adapting.
d. The mutations occurred because mosquitoes needed to survive.

16. The chemical deltamethrin (DM) is another insecticide that is used to kill mosquitoes. Not surprisingly, mosquito populations sprayed with DM for several years evolve to become resistant. However, if DM spraying is stopped for a few years, the population will lose its resistance. What is the most likely description of how this occurs?

a. Mosquitoes do not need to be resistant to DM when it is not present.
b. The immune systems of mosquitoes not exposed to DM gradually lose their ability to cope with DM.
c. When DM is not present, mosquitoes that are resistant to DM do not survive as well as mosquitoes that are not resistant to DM
d. This was purely a random event.



Cheetahs are large cats that hunt gazelles on the African savanna. Cheetahs are famous for being the fastest land animal: they can sprint up to 75 miles an hour. Cheetahs are also known for having spots that help them hide in bushes or tall grass. This, of course, is useful for sneaking up on gazelles. Cheetahs evolved their spots and incredible speed from ancestors that did not have spots or run as fast.

17. What is the best description of the process by which cheetahs evolved their distinctive spots?

a. Individuals adapted to their environment so they could stalk gazelles.
b. Cheetahs that did not blend into the environment could not catch gazelles and starved.
c. The shifting sunlight and shadows of the savanna changed the fur of cheetahs.
d. Changes like this depend on many factors, so it is impossible to answer.

18. The strength, endurance, and running speed of all animals is improved by exercise. This is true for human athletes and for cheetahs. What role did exercise play in the evolution of cheetahs?

a. Cheetahs evolved to be fast because they ran hard.
b. Cheetahs evolved to be fast, in part, because they ran hard, but also for other reasons.
c. Cheetahs evolved to be fast entirely because of other reasons.

19. Consider a cheetah that grew up during the time cheetahs were evolving to run fast. How did this cheetah's running speed likely compare to its parents?

a. The cheetah was probably a little faster than its parents.
b. The cheetah was probably about as fast as its parents.
c. The cheetah was probably a little slower than its parents.
d. Running speed is affected by so many factors, it is impossible to predict.

20. Think about the first cheetah to have spots like modern cheetahs. Why did this cheetah develop these spots?

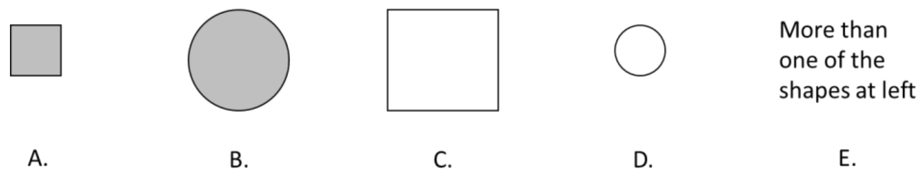
a. The cheetah needed better camouflage to hide in tall grass.
b. The cheetah was lucky a genetic error happened to give it these spots.
c. The cheetah's parents spent a lot of time hiding in tall grass and bushes.
d. The cheetah spent a lot of time hiding in tall grass and bushes.

APPENDIX B

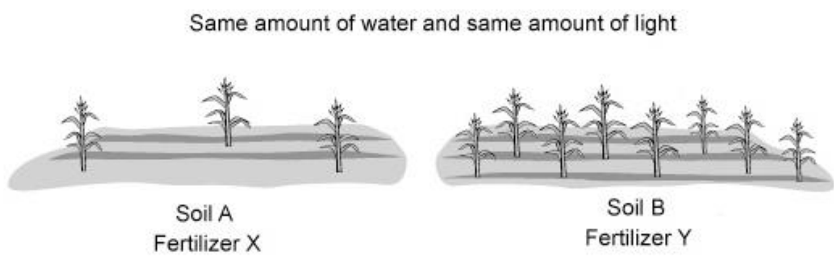
MONTANA STATE UNIVERSITY FORMAL REASONING TEST (MSU-FORT)

Montana State University Formal Reasoning Test (MSU-FORT)

21. Which of the shapes below violates the rule: “If it is white, it is round”?



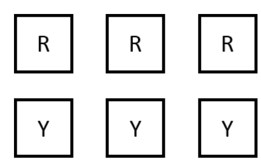
22. Hannah wants to know which type of soil is best for growing corn. She also wants to know which type of fertilizer is best. She performs an experiment using two types of soil (A and B) and two types of fertilizer (X and Y). The figure below shows what her corn looks like at the end of the summer:



What can Hannah conclude from this experiment?

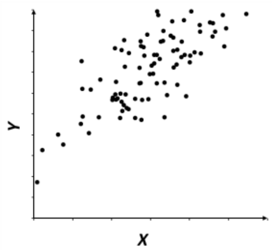
a. Soil B is best for growing her corn.
b. Fertilizer Y is best for growing her corn.
c. Soil B is best for growing her corn, and Fertilizer Y is best for growing her corn.
d. It is not possible to conclude which soil is best for growing her corn or which fertilizer is best for growing her corn.

23. Six square pieces of wood are put into a cloth bag and mixed about. The six pieces are identical in size and shape, however, three pieces are red and three are yellow. Suppose someone reaches into the bag (without looking) and pulls out one piece.



What are the chances the piece is red?

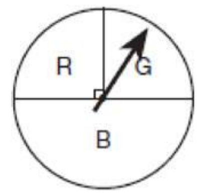
a. 1 chance out of 6
b. 1 chance out of 3
c. 1 chance out of 2
d. 1 chance out of 1
e. Cannot be determined



24. The graph on the left shows Variable X plotted against Variable Y. Does it appear that small values of X are associated with small values of Y, and that large values of X are associated with large values of Y?

a. Yes
b. No
c. You can't tell from this graph.

25. The spinner, shown at right, is spun twice. What is the probability the arrow will land in section **G** the first time and in section **B** the second time?



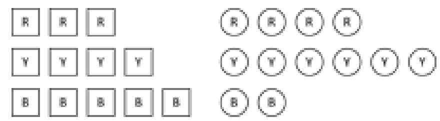
a. 1 / 9
b. 1 / 8
c. 1 / 4
d. 3 / 4
e. None of the above

26. Some children with disabilities are unable to communicate by talking, signing, or writing. A therapist believes he can help such a girl communicate by assisting her use a keyboard. He supports the girl's arm and uses subtle cues from the girl to bring the girl's fingers to keys on the keyboard. This appears to allow the girl to communicate for the first time in her life. A doctor, however, is skeptical. The doctor suggests the therapist may be *unconsciously* guiding the girl's hands to the keys, and that the messages are not from the girl at all.

How could you test whether the doctor is right?

a. Ask the therapist whether he really is selecting the letters.
b. Ask the girl if she really is selecting the letters on the keyboard.
c. Ask the girl a question only she knows the answer to.
d. Ask the girl a question you and she know the answer to, but the therapist does not.
e. There is no practical way to test whether the messages are coming from the girl.

27. Twenty-four blocks of wood are placed in a bag. Some are circles, and some are squares (see diagram). Some are red (R), some are yellow (Y), and some are blue (B). Someone reaches into the bag and pulls out one block without looking at the color or feeling the shape.



What is the chance the piece is a red round block or a blue round block?

a. 1 chance out of 2
b. 1 chance out of 3
c. 1 chance out of 4
d. 1 chance out of 6
e. None of the above

28. Assume the following two statements are true:

If someone eats candy every day, they are at risk of developing diabetes.
Alice does not eat candy every day.

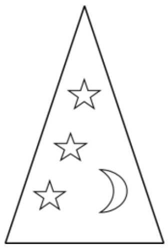
Given this information, is Alice at risk of developing diabetes?

a. Yes. b. No. c. Uncertain.

29. Aspen Valley is a valley in the Rocky Mountains that has a large population of elk. One year, a pack of wolves migrates to Aspen Valley. The wolves thrive. The elk do not do so well. In the two years after the wolves arrive, the number of elk in Aspen Valley declines by 50%. A biologist wants to know why. An obvious explanation is the wolves ate the elk. There is, however, another possible explanation. Aspen Valley had two harsh winters in a row, and deep snow may have killed young, old, or weak elk. Biologists have been counting elk in three nearby valleys. These valleys do not have wolves but did have the same harsh winters as Aspen Valley. Records show the number of elk in the three other valleys declined by 50%.

What is the simplest interpretation for these observations?

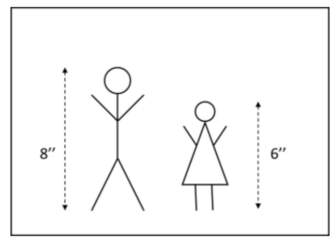
a. Wolves are responsible for the decline in elk numbers in Aspen Valley, and harsh winters are responsible for the decline in elk numbers in the other valleys.
b. Harsh winters are responsible for the decline in elk numbers in all four valleys.
c. Wolves are responsible for the decline in elk numbers in all four valleys.
d. Elk left Aspen Valley and moved to the other three valleys that did not have wolves.
e. The biologists need to watch the wolves and elk to see what is happening.



30. A girl is sewing a wizard's costume for Halloween. The diagram on the left is a scale-drawing of the pattern she is using to make a hat. The pattern is too small for the girl's head, so she increases the width of the hat by one inch. She also increases the height of the hat by one inch.

How will this affect the shape of the hat?
a. The hat will be more pointed.
b. The hat will be less pointed.
c. The hat will be the same shape.
d. Not enough information is provided to answer.

31. Here is a photograph of a father and a daughter. In the photograph, the father is 8 inches tall and the daughter is 6 inches tall. The family decides to enlarge the picture so that the father will be 12 inches tall.



How tall will the daughter be in the enlarged picture?
a. 6 inches
b. 8 inches
c. 9 inches
d. 10 inches
e. None of the above

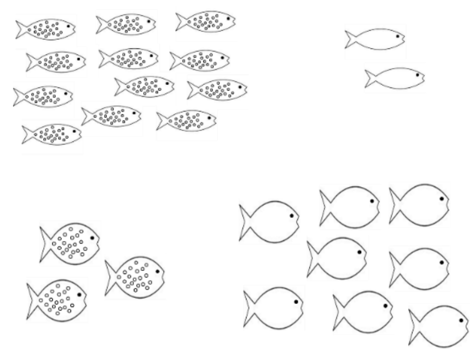
32. Homing pigeons are famous for being able to fly home after being released far from their home. Vivian wants to know how pigeons find their way home. In particular, she wants to know if pigeons have an internal compass and use Earth's magnetic field to navigate. Vivian tests this hypothesis by gluing a small bar magnet to the backs of 16 homing pigeons. She thinks these magnets are strong enough to prevent the pigeons from sensing Earth's natural magnetic field, but doesn't know for sure. Another potential problem is that the magnets weigh 2 grams. Vivian hopes this is small enough they won't interfere with the pigeons' ability to fly, but isn't sure. She, therefore, glues a small aluminum bar that weighs 2 grams to the backs of 16 other pigeons. Vivian drives all of the pigeons 100 miles from her home and releases them. The table below lists four possible results for the experiment.

Considering only the information provided here, what result would be hardest for Vivian to explain?

	Pigeons with:	
	2 gram magnet	2 gram aluminum bar
a.	Return	Return
b.	Do not return	Do not return
c.	Do not return	Return
d.	Return	Do not return

e. This experiment has too many potential problems for the results to be interpreted.

33. Katherine catches 25 fish. The fish all belong to the same species, but are not all the same. Two traits vary: the shape of the fish, and whether they have spots.



Does there seem to be an association between the shape of the fish and whether it has spots?

a. Yes.
b. No.
c. You can't tell from this sample.

34. Assume the following two statements are true:

If Iris works hard, she will finish her homework before dinner.
Iris did not finish her homework before dinner.

What can we conclude from this information?

a. Iris worked hard at her homework.
b. Iris did not work hard at her homework.
c. We cannot tell if Iris worked hard or not.

35. Brendan wants to know how trees move water from their roots up to their leaves. One potential explanation is that cells in the roots have molecular pumps that push water upwards. Another explanation is that leaves have molecular pumps that suck water upwards. Brendan cuts the roots off several small trees and places the trees in buckets of water containing red food coloring. An hour later, he observes that the water containing the red food coloring has risen to the top of these trees.

What can Brendan conclude?

a. Molecular pumps in roots are pushing water to the top of his trees.
b. Molecular pumps in leaves are sucking water to the top of his trees.
c. Molecular pumps in roots are not responsible for moving water up his trees.
d. He cannot conclude much because the trees in his experiment did not have roots.
e. He cannot conclude much because his experiment did not have a control.

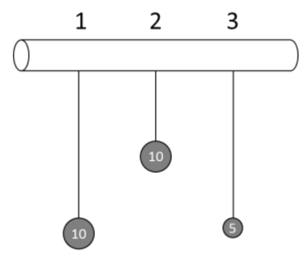
36. A medical researcher is trying to develop a blood test to identify patients that have a high risk of heart attack. He believes a certain enzyme in patients' blood might be a warning sign of a heart attack. He tests 70 patients for this enzyme and monitors them for five years. The table below shows whether or not they had a heart attack.

Did patient have heart attack?		
	No	Yes
Enzyme present	36	4
Enzyme absent	27	3

Does there appear to be an association between the enzyme and heart attacks?

a. Yes
b. No
c. You can't tell from the data.

37. The drawing on the right shows three strings hanging from a bar. Each string has a metal weight at the end that weighs 5 or 10 ounces. The strings (and attached weights) can be swung back and forth, and the time it takes for the weight to swing back and forth can be measured. Suppose you want to find out whether the <u>length</u> of the string has an effect on how long it takes for the string to swing back and forth.



Which string(s) would you use to find out?

a. Any string   b. All 3 strings   c. 1 and 2   d. 1 and 3.   e. 2 and 3

38. Emma and Sarah are making lemonade by stirring powered lemonade mix into water. The table at right shows how much lemonade mix and how much water each girl uses.

	Lemonade mix	Water
Emma	1 table spoon	2 cups
Sarah	2 table spoons	3 cups

Whose lemonade will have a stronger flavor?

a. Emma's
b. Sarah's
c. Both will taste the same.
d. It depends on how strong the lemonade mix is.

39. Assume the following two statements are true:

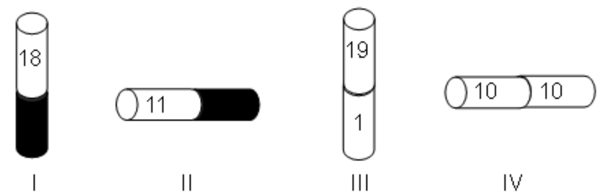
If John's mother has surgery on Friday, John will be at the hospital.
John was at the hospital on Friday.

Given this information, did John's mother have surgery on Friday? a. Yes b. No. c. Uncertain

40. Twenty fruit flies are placed in each of four glass tubes. The tubes are sealed. Tubes I and II are partially covered with black paper; Tubes III and IV are not covered. The tubes are placed as shown. Then they are exposed to red light for five minutes. The number of flies in the uncovered part of each tube is shown in the drawing.

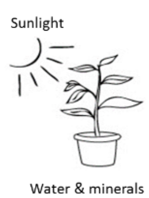
This experiment shows that flies move towards or away from:

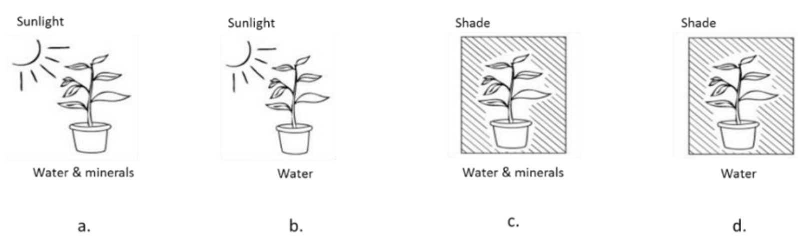
a. Red light but not gravity.
b. Gravity but not red light.
c. Both red light and gravity.
d. Neither red light nor gravity.
e. It cannot be determined.



All tubes are exposed to red light

41. A student wants to find out if a particular kind of plant grows better in the sun or in the shade. She has two identical potted plants. She gives one plant water and minerals and places the plant in sunlight (see figure at left). Which of the following conditions should she use for the second plant to study the effect of sunlight on plants?

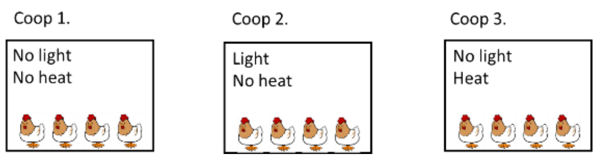




42. Kate has noticed her chickens lay fewer eggs during the winter than in the summer and wants to know why. She has two ideas:

Her chickens are laying fewer eggs because it is cold outside.
Her chickens are laying fewer eggs because it is dark outside.

She decides to test these potential explanations the next winter in three chicken coops. The first coop is the type of coop she has always used: it has no light or heat. The second coop has a light bulb but no heat. The third coop has a heater but no light.



What will Kate observe if temperature affects how often chickens lay eggs and light does not?

a. The chickens with light but not heat (Coop 2) will lay more eggs than the other coops.
b. The chickens with heat but not light (Coop 3) will lay more eggs than the other coops.
c. The chickens with light or heat (Coops 2 & 3) will lay more eggs than chickens in Coop 1.
d. All the chickens will lay the same number of eggs.
e. There is no way to predict what will happen.

APPENDIX C

SCIENCE MOTIVATION QUESTIONNAIRE II (SMQ-II)

Science Motivation Questionnaire II (SMQ-II)

To better understand what you think and how you feel about your biology courses, please respond to each of the following statements from the perspective of “When I am in this biology course…’’

	Never	a	b	c	d	e	Always
43. The biology I learn is relevant to my life.		a	b	c	d	e	
44. I like to do better than other students on tests.		a	b	c	d	e	
45. Learning biology is interesting		a	b	c	d	e	
46. Getting a good biology grade is important to me.		a	b	c	d	e	
47. I put enough effort into learning biology.		a	b	c	d	e	
48. I use strategies to learn biology well.		a	b	c	d	e	
49. Learning biology will help me get a good job.		a	b	c	d	e	
50. It is important that I get an “A” in biology.		a	b	c	d	e	
51. I am confident I will do well on biology tests.		a	b	c	d	e	
52. Knowing biology will give me a career advantage.		a	b	c	d	e	
53. I spend a lot of time learning biology.		a	b	c	d	e	
54. Learning biology makes my life more meaningful.		a	b	c	d	e	
55. Understanding biology will benefit me in my career.		a	b	c	d	e	
56. I am confident I will do well on biology labs.		a	b	c	d	e	
57. I believe I can master biology knowledge and skills.		a	b	c	d	e	
58. I prepare well for biology tests and labs.		a	b	c	d	e	
59. I am curious about discoveries in biology.		a	b	c	d	e	
60. I believe I can earn a grade of “A” in biology.		a	b	c	d	e	
61. I enjoy learning biology.		a	b	c	d	e	
62. I think about the grade I will get in biology.		a	b	c	d	e	
63. I am sure I can understand biology.		a	b	c	d	e	
64. I study hard to learn biology.		a	b	c	d	e	
65. My career will involve biology.		a	b	c	d	e	
66. Scoring high on biology tests and labs matters to me.		a	b	c	d	e	
67. I will use biology problem-solving skills in my career.		a	b	c	d	e	

APPENDIX D

CARL WIEMAN SCIENCE EDUCATION INITIATIVE

TEACHING PRACTICES INVENTORY

(CWSEI_TPI)

Carl Wieman Science Education Initiative Teaching Practices Inventory (CWSEI_TPI)

Please complete this survey on teaching practices. As you will see, the list is comprehensive, and we realize that any one course would most likely use only a subset of these practices. The survey was developed by Wieman and Gilbert (CBE-Life Sciences Education, 2014, 13(3), 552-569).

Q.I

1. What course information do you provide to students via hard copy or course web page? Check all that apply in your course.
   - ☐ List of topics to be covered. [1]
   - ☐ List of topic-specific competencies (skills, expertise, ...) students should achieve. (What student should be able to do.) [2]
   - ☐ List of competencies that are not topic related (critical thinking, problem solving, ...). [3]
   - ☐ Affective goals - changing students’ attitudes and beliefs (interest, motivation, relevance, beliefs about competencies, how to master the material). [4]
   - ☐ Other (please specify below). [5]

   If you selected Other above, please specify here. ___________________ [6]

Q.II

1. What supporting materials do you provide to students? Check all that apply in your course.
   - ☐ Student wikis or discussion boards with little or no contribution from you. [1]
   - ☐ Student wikis or discussion boards with little significant contribution from you or Teaching Assistant. [2]
   - ☐ Solutions to homework assignments. [3]
   - ☐ Worked examples (text, pen-cast, or another format). [4]
   - ☐ Practice or previous year's exam. [5]
   - ☐ Animations, video clips, or simulations related to the course. [6]
   - ☐ Lecture notes or course PowerPoint presentations (partial, skeleton, or complete). [7]
   - ☐ Other instructor selected notes or supporting materials, pen-casts, etc. [8]
   - ☐ Articles from scientific literature. [9]
   - ☐ Other (please specify below). [10]

   If you selected Other above, please specify here. ___________________ [11]

Q.III

1. What is the average number of times per class you pause to ask for questions?
2. What is the average number of times per class you have small group discussions or problem solving?
3. What is the average number of times per class you show demonstrations, simulations, or video clips?
4. What is the average number of times per class you show demonstrations, simulations, or video where students first record predicted behavior and then afterward explicitly compare observations with predictions?
5. What is the average number of discussions per term on why materials are useful and/or interesting from students' perspective?
6. Comments on above questions (if any). ____

   ____

7. Which of the following activities do you use?
   Check all that apply in your course.
   □ Students asked to read/view material on upcoming class session. [1]
   □ Students read/view material on upcoming class session and complete assignments or quizzes on it shortly before class or at beginning of class. [2]
   □ Reflective activity at end of class, e.g. “one-minute paper” or similar (students briefly answering questions, reflecting on lecture and/or their learning, etc.). [3]
   □ Student presentations (verbal or poster) [4]
8. What fraction of a typical class period do you spend lecturing (presenting content, deriving mathematical results, presenting a problem solution, …)?
   □ 0-20% [1]
   □ 20-40% [2]
   □ 40-60% [3]
   □ 60-80% [4]
   □ 80-100% [5]
9. Considering the time spent on the major topics, approximately what fraction was spent on the process by which the theory/model/concept was developed?
   □ 0-10% [1]
   □ 11-25% [2]
   □ more than 25% [3]
10. If a personal response system (PRS) is used to collect responses from all students IN REAL TIME IN CLASS, what method is used?
    Check all that apply in your course.
    □ electronic (“clickers”) with student identifier [1]
    □ electronic anonymous [2]
    □ colored cards [3]
    □ raising hands [4]
    □ written student responses that are collected and reviewed in real time [5]
    □ Other (please specify below). [6]
    If you selected Other above, please specify here. __________________ [7]
11. If you use a PRS, how many PRS questions do you pose per class that are followed by student-student discussion? ______
12. If you use a PRS, how many times per class do you use PRS as a quiz device (counts for marks and no student discussion)? ______

Q.IV

1. Which of the following assignments do you use?
   Check all that apply in your course.
   □ Problem sets/homework assigned or suggested but did not contribute to course grade [1]
   □ Problem sets/homework assigned and contributed to course grade at intervals of 2 weeks or less [2]
   □ Paper or project (an assignment taking longer than two weeks and involving some degree of student control in choice of topic or design) [3]
   □ Encouragement and facilitation for students to work collaboratively on their assignments [4]
   □ Explicit group assignments [5]
   □ Other (please specify below). [6]
   If you selected Other above, please specify here. __________________ [7]

Q.V

1. Which of the following feedback and testing do you use?
   Check all that apply in your course.
   ☐ Midterm course evaluation [1]
   ☐ Repeated online or paper feedback or via some other collection means such as clickers [2]
   ☐ Other (please specify below). [3]
   If you selected Other above, please specify here. ___________________ [4]
2. What kind of feedback and testing occurs in your course?
   Check all that apply in your course.
   ☐ Assignments with feedback before grading or with opportunity to redo work to improve grade. [1]
   ☐ Students see graded assignments. [2]
   ☐ Students see assignment answer key and/or grading rubric. [3]
   ☐ Students see graded midterm exam(s). [4]
   ☐ Students see midterm exam(s) answer key(s). [5]
   ☐ Students explicitly encouraged to meet individually with you. [6]
   ☐ Other (please specify below). [7]
   If you selected Other above, please specify here. ___________________ [8]
3. What is the number of midterm exams in your course? ______
4. What is the approximate percentage (%) of exam points from questions that require students to explain reasoning? ______
5. What is the approximate percentage (%) breakdown of the course grade?
   (Total must add up to 100.)
   ______ Final exam [1]
   ______ Midterm exam(s) [2]
   ______ Homework assignments [3]
   ______ Paper(s) or project(s) [4]
   ______ In-class activities [5]
   ______ In-class quizzes [6]
   ______ Online quizzes [7]
   ______ Participation [8]
   ______ Lab component [9]
   ______ Other (please specify below) [10]
   If you selected Other above, please specify here. ___________________ [11]

Q.VI

1. Other practices not previously mentioned.
   Check all that apply in your course.
   ☐ Assessment given at beginning of course to assess background knowledge. [1]
   ☐ Use of instructor-independent pre-test/post-tests (e.g. concept inventory) to measure learning. [2]
   ☐ Use of a consistent measure of learning that is repeated in multiple offerings of the course to compare learning. [3]
   ☐ Use of pre-post survey of student interest and/or perceptions about the subject. [4]
   ☐ Opportunities for students' self-evaluation of learning. [5]
   ☐ Students provided with opportunities to have some control over their learning, such as choice of topics for course, paper, or project, choice of assessment methods, etc. [6]

☐ New teaching methods or materials were tried along with measurements to determine their impact on student learning. [7]

Q.VII

1. Training and guidance of Teaching Assistants.
   Check all that apply in your course.
   ☐ No TAs for course. [1]
   ☐ TAs must satisfy English language skills criteria. [2]
   ☐ TAs receive ½ day or more of training in teaching. [3]
   ☐ Use of pre-post survey of student interest and/or perceptions about the subject. [4]
   ☐ There are Instructor-TA meetings every two weeks or more frequently where student learning and difficulties, and the teaching of upcoming material are discussed. [5]
   ☐ TAs are undergraduates. [6]
   ☐ TAs are graduate students. [7]
   ☐ Other (please specify below). [8]
   If you selected Other above, please specify here. ___________________ [9]

Q.VIII

1. What kind of collaboration or sharing strategies do you use in your teaching?
   Check all that apply in your course.
   ☐ Used or adapted materials provided by colleague(s). [1]
   ☐ Used "Departmental" course materials that all instructors of this course are expected to use. [2]
2. How frequently do you do the following?
   Discussed how to teach the course with colleague(s). [1]

	1	2	3	4	5
Never versus Very Frequently	☐	☐	☐	☐	☐

Read literature about teaching and learning relevant to this course. [2]

	1	2	3	4	5
Never versus Very Frequently	☐	☐	☐	☐	☐

Sat in on colleague's class (any class) to get/share ideas for teaching. [3]

	1	2	3	4	5
Never versus Very Frequently	☐	☐	☐	☐	☐

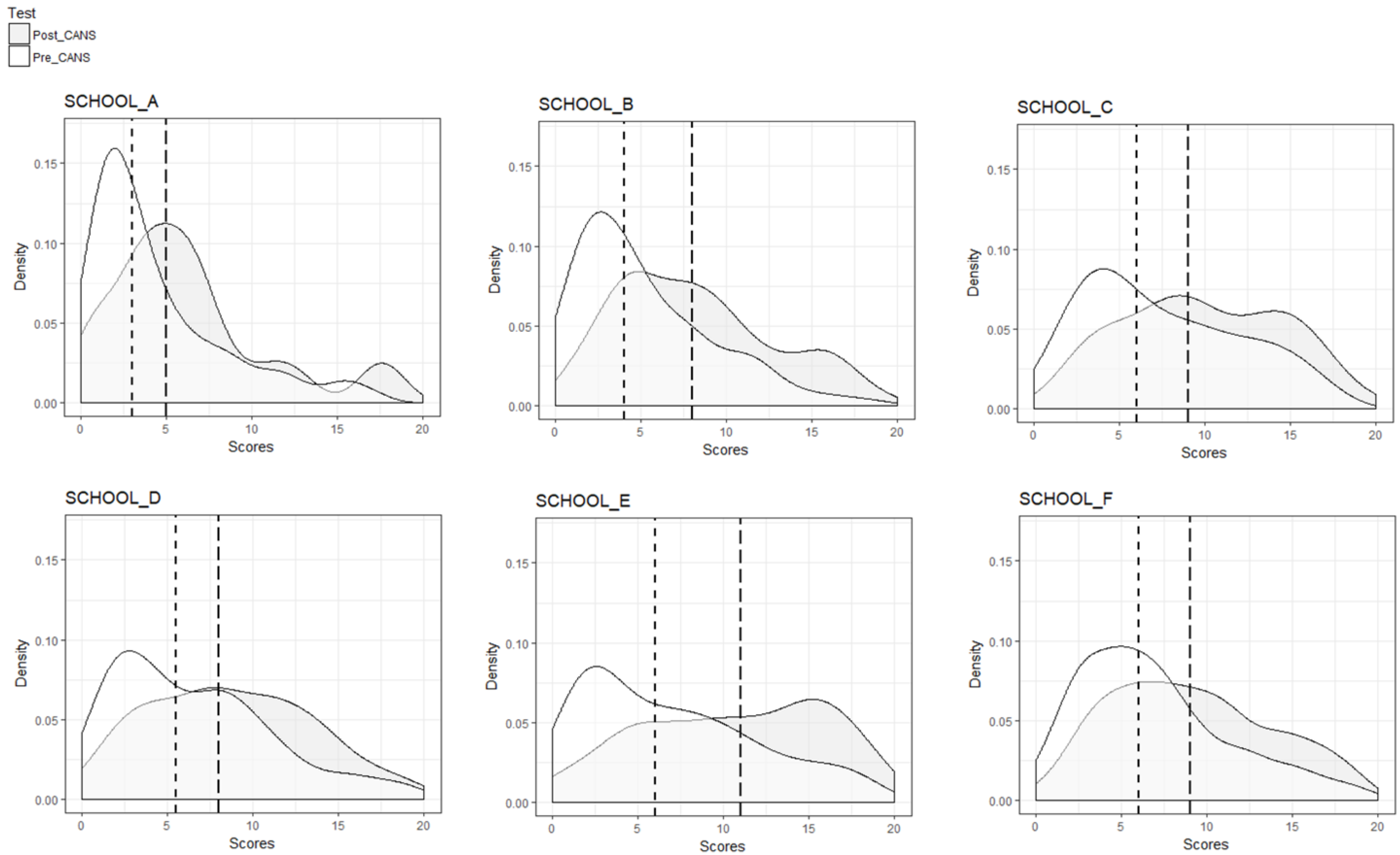
Q.IX

1. General (open-ended comments). Please write any other comments here. If this inventory has not captured an important aspect of your teaching of this course, or you feel you need to explain any of your above answers please describe it here. ______________________________
____
____

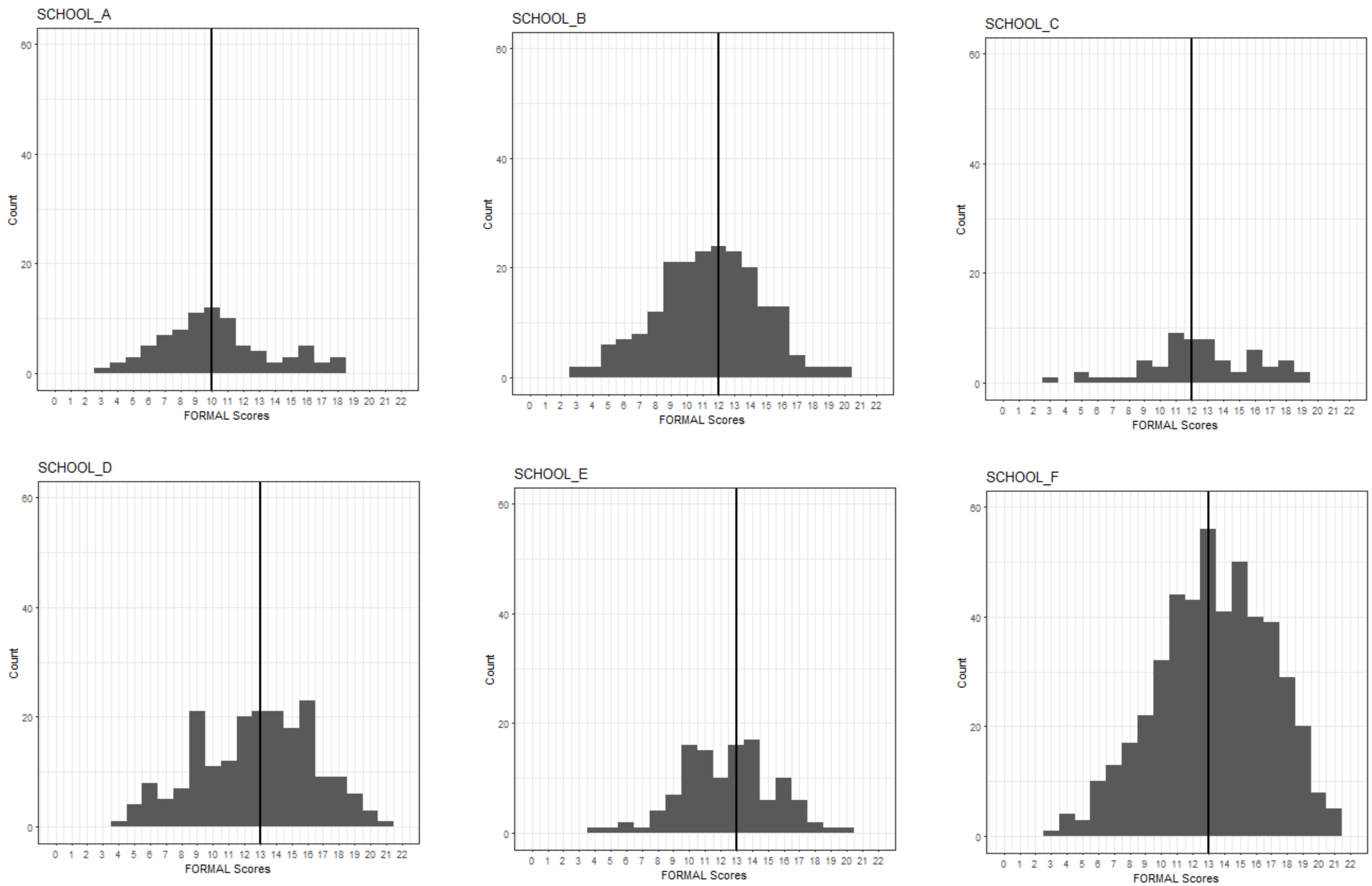
APPENDIX E

SUMMARY OF RESULTS FOR PRE- & POST-CANS,

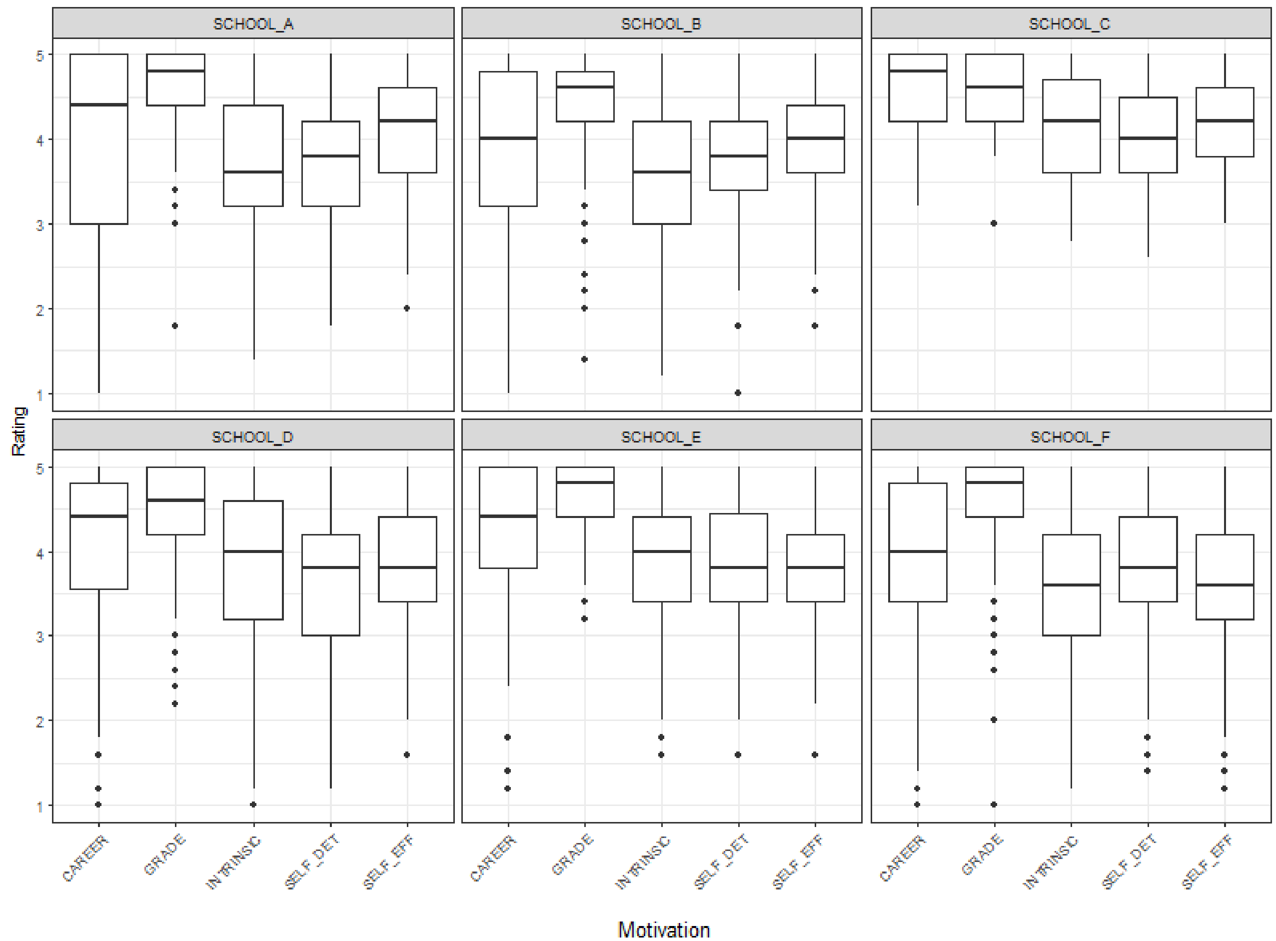
MSU-FORT, AND SMQ-II BY SCHOOL



Plot of density curves for pre- and post-instruction CANS scores with pre-CANS (dash line) and post-CANS (longdash line) median scores by school



Histogram of formal reasoning scores with median (solid line) score by school



Boxplot of ratings for motivation variables with median (solid line) scores by school