

5-31-2016

Evaluation of Students' Interest, Effort Beliefs, and Self-Efficacy in General Chemistry

Brent Ferrell

Follow this and additional works at: <http://digscholarship.unco.edu/dissertations>

Recommended Citation

Ferrell, Brent, "Evaluation of Students' Interest, Effort Beliefs, and Self-Efficacy in General Chemistry" (2016). *Dissertations*. 345.
<http://digscholarship.unco.edu/dissertations/345>

This Text is brought to you for free and open access by the Student Research at Scholarship & Creative Works @ Digital UNC. It has been accepted for inclusion in Dissertations by an authorized administrator of Scholarship & Creative Works @ Digital UNC. For more information, please contact Jane.Monson@unco.edu.

UNIVERSITY OF NORTHERN COLORADO

Greeley, Colorado

The Graduate School

EVALUATION OF STUDENTS' INTEREST, EFFORT BELIEFS,
AND SELF-EFFICACY IN GENERAL CHEMISTRY

A Dissertation Submitted in Partial Fulfillment
of the Requirements for the Degree of
Doctor of Philosophy

Brent Ferrell

College of Natural and Health Sciences
Department of Chemistry and Biochemistry
Chemical Education

July 2016

This Dissertation by: Brent Ferrell

Entitled: *Evaluation Of Students' Interest, Effort Beliefs, And Self-Efficacy In General Chemistry*

has been approved as meeting the requirement for the Degree of Doctor of Philosophy in the College of Natural and Health Sciences in the Department of Chemistry and Biochemistry, Program of Chemical Education

Accepted by the Doctoral Committee

Jack Barbera, Ph.D., Co-Research Advisor

Kimberly A. O. Pacheco, Ph.D., Co-Research Advisor

Murielle Watzky, Ph.D., Co-Research Advisor

Richard Hyslop, Ph.D., Committee Member

Michael Phillips, Ph.D., Faculty Representative

Date of Dissertation Defense _____

Accepted by the Graduate School

Linda L. Black, Ed.D.
Associate Provost and Dean
Graduate School and International Admissions

ABSTRACT

Ferrell, Brent. *Evaluation Of Students' Interest, Effort Beliefs, And Self-Efficacy In General Chemistry*. Published Doctor of Philosophy dissertation. University of Northern Colorado, 2016.

The research described in this dissertation is outlined in three phases and is focused on the measurement, interrelationships, and understanding of self-efficacy, interest, and effort beliefs among general chemistry students. The primary drive behind this research was to provide measurement tools to chemistry education researchers and practitioners as a way to evaluate novel and alternative teaching strategies and interventions.

The first phase of this study involved gathering evidence for validity and reliability of four previously published scales that measured initial interest, maintained interest, self-efficacy, and effort beliefs. These scales were taken from other disciplines, with the exception of the self-efficacy scale, and modified to fit into a general chemistry context. This phase of the study involved both quantitative and qualitative methods. On the quantitative side, confirmatory factor analysis was used in a pilot study ($n_1 = 373$, $n_2 = 294$) and a cross-validation study ($n = 1,160$) to evaluate how well the items in each scale described a single construct among general chemistry students. In addition, the changes in students' scores across the semester were calculated for a sub-sample of each of the full samples in the pilot and cross-validation studies. The qualitative thread

included interviewing students from the target population to assess the readability and interpretation of each item from all of the scales. The results of both the quantitative and qualitative analyses were reviewed concurrently to remove problematic items. The four scales were modified by removing a total of five items, resulting in improved model fit and better understanding among students.

The second phase of this study built on the first phase by utilizing the modified scales to test the connections between self-efficacy, interest, and effort beliefs as well as their relation to course performance. A total of 143 participants from first-semester general chemistry were included in the analyses, which utilized path analysis, multiple regression, and MANCOVA. Data were collected twice during the semester – once during the first week (time 1) and again during the thirteenth week (time 2). The results revealed that time 2 measures were superior at predicting course grade than time 1 measures. The final model accounted for 34% of the total variance in course grade.

The third phase of the study was entirely qualitative and focused on interviewing general chemistry students about the sources and influences of their effort beliefs. Since very little has been reported about effort beliefs, the objective here was to expand what is currently known about how college students acquire their beliefs about effort. A total of 21 students were interviewed over the course of three semesters. Two major sources of effort beliefs were reported by the interviewees – family influence and personal experiences. Most of the participants alluded to one of these, with a few participants mentioning both. By

understanding more about where effort beliefs originate, instructors can work toward implementing methods that will target and enhance their students' effort beliefs.

ACKNOWLEDGEMENTS

I owe a tremendous amount of gratitude to many people who have guided me throughout my educational journey. First and foremost, I want to thank my Lord and Savior, Jesus Christ, for His guidance and blessings along the way. I would not be writing this today without Him in my life.

Kathy, thank you for supporting this endeavor and achievement. I could not have succeeded at this level without your giving spirit. It has meant so much to have a devoted partner to bless me and encourage me when I needed it most.

Dr. Barbera, thank you for always keeping your door open and for your clear, concise direction. You have truly been a role model for me not only in how you conduct research and advise your group, but also in teaching. I greatly appreciate the space you gave me to work on my own, while always being available for any questions and concerns I had. I am proud to have been a part of your research group.

Dr. Phillips, thank you for making the world of motivation research more digestible for a chemist. I appreciate your patience in working with me and all of the meetings we had to hash out data and interpretations. And, thank you for challenging me to think and see the research problem from multiple angles.

To my other committee members, thank you for all of your guidance along the way. I appreciate the time you devoted to my project, your commitment to my success, and for all of your encouragement during my time at UNC.

Dr. Hutchinson and Dr. Teman, thank you both for all your time during the last 4 years, both in and out of the classroom. I appreciate your willingness to really think about and work through the statistical problems I was having. I have gained a wealth of knowledge in the field of applied statistics thanks to you two.

To my previous research advisors and educators, thank you for helping to pave the way for this accomplishment. My knowledge and drive for success is a product of your investment in my education and your encouragement as well as your example for what it means to be a scholar.

TABLE OF CONTENTS

CHAPTER I. INTRODUCTION	1
Statement of Problem	5
Purpose of Study	7
Research Questions	10
Significance of Study	11
Limitations	13
Definition of Terms	14
CHAPTER II. REVIEW OF LITERATURE	16
Teaching and Learning in College Chemistry	17
A Brief History of Academic Motivation Research	29
Self-Efficacy	37
Overview of social cognitive theory	37
Self-efficacy beliefs	41
Sources of self-efficacy	42
Constructs related to self-efficacy	46
Self-efficacy and academic achievement	50
Self-efficacy in college chemistry	52
Interest	55
Interest and education	61
Interest as a variable in chemistry	65
Effort Beliefs	69
Brief overview of attribution theory	69
Implicit theories of intelligence	70
Prior research and implications of effort beliefs	73
Connections Between Effort Beliefs, Interest, and Self-Efficacy	78
Social Psychological Interventions in Education	82
Psychometrics	89
Test dimensionality and factor analysis	91
Validity	94
Evidence for test content validity	96
Evidence for response process validity	96
Evidence for relationships with other variables	98
Evidence for criterion-related validity	98
Evidence for internal structure validity	101

Summary	101
Reliability	102
Existing Motivation Instruments and Scales	107
Summary.....	115
 CHAPTER III. METHODOLOGY	 118
Introduction	118
Research Questions	119
Protection of Human Subjects	119
Phase one: Adaptation of Scales for Use in General Chemistry	120
Research design	120
Scales	121
Participants	125
Initial survey data collection	125
Cross-validation survey data collection	126
Survey data analysis	127
Descriptives.....	127
Time 1 and time 2 measurements.....	127
Confirmatory factor analysis.....	129
Interview participants	132
Interview design and protocol	133
Interview data collection and analysis.....	133
Quantitative-Qualitative integration	134
Phase two: Classroom-Based Study	134
Research design	134
Participants	136
Scales and demographic form	136
Survey data collection	137
Performance data collection.....	137
Interventions.....	137
Pre/post data analysis.....	140
Intervention data analysis	145
Phase three: Effort Beliefs Interviews	149
Participants	149
Interview design and protocol	150
Interview data collection and analysis.....	151
 CHAPTER IV. ARTICLE 1: ANALYSIS OF STUDENTS' SELF-EFFICACY INTEREST, AND EFFORT BELIEFS IN GENERAL CHEMISTRY	 153
Abstract.....	153
Introduction	155
Methods	171
Results.....	183
Discussion.....	203

CHAPTER V. ARTICLE 2: ANALYSIS OF STUDENTS' SELF-EFFICACY INTEREST, AND EFFORT BELIEFS IN GENERAL CHEMISTRY	214
Abstract.....	214
Introduction	215
Methods	223
Results	229
Discussion.....	238
CHAPTER VI. CONCLUSIONS, IMPLICATIONS, AND FUTURE RESEARCH	247
Conclusions	247
Implications	262
Future Research	265
REFERENCES	267
APPENDIX A. INSTITUTIONAL REVIEW BOARD APPROVAL AND INFORMED CONSENT FORMS	312
APPENDIX B. ANNOUNCEMENTS FOR PARTICIPATION.....	318
APPENDIX C. SUPPORTING INFORMATION FOR QUANTITATIVE STUDIES	323
APPENDIX D. TIME TABLE AND SUMMARY FOR ALL INTERVENTIONS.....	334
APPENDIX E. COPYRIGHT PERMISSION FOR ARTICLE 1	347

LIST OF TABLES

1.	CFA fit indices and reliability estimates of preliminary and revised scales at time 1 for the initial sample	190
2.	CFA fit indices and reliability estimates of preliminary and revised scales at time 2 for the initial sample	193
3.	CFA fit indices and reliability estimates for initial and cross-validation samples.....	195
4.	Mean values of scores for initial and maintained interest at times 1 and 2 for all students from the initial sample.....	197
5.	Mean values of scores for self-efficacy and effort beliefs at times 1 and 2 for all students from both the initial sample and cross-validation sub-sample	199
6.	Mean scores and differences between chemistry majors and non-science majors on interest and self-efficacy scales for the initial sample	201
7.	Mean values of self-efficacy scores by major at time 1 and time 2 from the initial sample	203
8.	Descriptive statistics for each measured scale	230
9.	Pearson correlations of all scales and final course grade.....	231
10.	Fit statistics for each model tested.....	233

LIST OF FIGURES

1.	Representation of the contemporary viewpoint of validity evidences for testing purposes	96
2.	Path diagram of models 1 and 2	234
3.	Path diagram of models 3-5	237

CHAPTER I

INTRODUCTION

Motivation, by definition, means “the reason or reasons one has for acting or behaving in a particular way” (Oxford Dictionaries Online). Over the last 40 years, the emphasis on motivation research has shifted from behavioral aspects such as drive and reinforcement to beliefs, goals, and expectations (Wigfield & Eccles, 2002b). The preliminary turn in motivation research occurred in the years around 1960 when psychologists moved away from mechanism toward cognition as a major source for explaining behavior (Weiner, 1990a). No longer were human behaviors simply regarded as resulting from either a reward or punishment. But rather, the way a reward or punishment is perceived by an individual is considered to be a principal key in understanding the behavioral outcome of that individual.

Motivation has been the focus of thousands of research articles in educational literature, with the bulk of those being published over the last 30 years. Many theories exist on the development, persistence, and actualization of motivation in the classroom. Some of the most prominent theories include: self-determination theory (Ryan & Deci, 2000b), expectancy-value theory (Wigfield & Eccles, 2002b), self-efficacy theory (Bandura, 1977), attribution theory (Weiner,

1985), interest theory (Schiefele, 1991), and goal theory (Elliott & Dweck, 1988). Although each of these theories differs in one way or another, one common thread stitched among all of them is the compelling importance of student motivation in the classroom.

In order for the motivation of students to be properly assessed, there must be quality measurement tools in place. Ensuring that the specific psychological variable intended to be measured is actually being measured may be the most difficult task in this line of research (Hinkin, 1995). The tools used by researchers to measure variables related to a psychological state, such as motivation, are referred to as instruments or scales. Many instruments and scales have been developed to measure variables tied to academic motivation such as: self-efficacy (Baldwin, Ebert-May, & Burns, 1999; Dalgety, Coll, & Jones, 2003; Dowling, 1978; Glynn, Taasoobshirazi, & Brickman, 2009; Gungor, Eryilmaz, & Fakioglu, 2007; Lawson, Banks, & Logvin, 2007; Midgley et al., 2000; Pintrich, Smith, García, & McKeachie, 1993; Tuan, Chin, & Shieh, 2005; Uzuntiryaki & Aydin, 2009), interest (Adams et al., 2006; Gungor et al., 2007; Harackiewicz, Durik, Barron, Linnenbrink-Garcia, & Tauer, 2008; Linnenbrink-Garcia et al., 2010), positive learning strategies (Blackwell, Trzesniewski, & Dweck, 2007), self-concept (Gungor et al., 2007; Marsh, 1992; Piers, 2002), goal orientation (Midgley et al., 2000; Pintrich et al., 1993; Tuan et al., 2005), self-determination (Glynn et al., 2009; Vallerand et al., 1992), and effort beliefs (Blackwell, 2002; Sorich & Dweck, 1997). These instruments and scales were either developed for general use in academia (Blackwell, 2002; Blackwell et al., 2007; Marsh, 1992;

Midgley et al., 2000; Piers, 2002; Pintrich et al., 1993; Vallerand et al., 1992) or for use in a specific subject area (Adams et al., 2006; Baldwin et al., 1999; Dalgety et al., 2003; Dowling, 1978; Glynn et al., 2009; Gungor et al., 2007; Harackiewicz et al., 2008; Lawson et al., 2007; Linnenbrink-Garcia et al., 2010; Uzuntiryaki & Aydin, 2009). Although useful during the primary and early secondary school years, instruments designed for general academia lack the precision necessary for older students, because students' beliefs about themselves tend to become more differentiated with age (Harter, 1998; Wigfield & Eccles, 2002a). For example, a student who once loved all subjects in school may find that she is more competent in math than language arts. As a result, her interest could shift toward math, and her responses concerning the subjects of math and language arts on a survey would be different. Despite this, the specificity of the survey used depends on the research question(s) being asked. Vallerand et al. (1992) developed the Academic Motivation Scale intended for college students. The items focus on one's motivation for attending college, as a life choice, and less about one's motivation in a given domain or subject area. Hence, it would be appropriate to use a survey such as this to determine why a sample of college students chose to attend college. While meaningful, this question leaves a lot of ground to be covered, such as: what factors are important for retaining students in science majors? What teaching practices are most appropriate for fostering positive learning strategies and promoting confidence in one's ability? Questions like these are best answered in the context of a specific discipline.

Recently, there has been a call for discipline-specific research on student motivation, particularly in the sciences. As stated in the 2012 Discipline-Based Education Report (DBER), students' dispositions and motivations to learn science and engineering are largely understudied and are of "central importance" (National Research Council, 2012). However, discipline-specific research on motivation is not possible without discipline-specific instruments to measure it. Researchers in chemical education have developed or modified several instruments designed to measure components of motivation in a chemistry classroom setting (Barbera, Adams, Wieman, & Perkins, 2008; Bauer, 2005, 2008; Dalgety & Salter, 2002; Uzuntiryaki & Aydin, 2009; Zusho, Pintrich, & Coppola, 2003). However, most of these instruments are long and cumbersome to administer. Moreover, some of the instruments are purported to measure certain constructs of motivation, but lack the necessary theoretical support. And finally, there are constructs related to motivation that have never been explored in a chemistry context.

The focus of this dissertation study is to contribute to the growing body of knowledge on motivation in introductory college chemistry settings. Specifically, the relationships among three motivational constructs (self-efficacy, interest, and effort beliefs) and their connection to course performance were investigated in several first-semester general chemistry sections. Thus, the guiding research question for this study is: How are self-efficacy, interest, and effort beliefs toward chemistry interconnected and to what extent do these predict course performance among general chemistry students? Once we understand more

about these phenomena, the chemistry education community can then begin to investigate how these aspects of motivation are impacted by various teaching practices.

Statement of Problem

There is growing concern over the declining rate of retention of undergraduate students from science, technology, engineering, and mathematics (STEM) disciplines in the United States (President's Council of Advisors on Science and Technology, 2012). It has been reported that the undergraduate rate of completion in STEM fields is lower than other fields for all ethnicities (National Research Council, 2012). Chemistry is a central requirement for most science fields as well as many health-related majors. The difficulty of chemistry, as evidenced by high D,F,W rates (Cooper & Pearson, 2012; McFate & Olmsted, 1999), poses a stumbling block for many students in science and health majors. Chemistry comes with many challenges for students including: solving algebraic expressions and proportions, understanding a complex chemical language, and interpreting atomic-level representations. In addition to the inherent challenges of the material, many college chemistry students are less than satisfied with the way it has been taught (Cooper, 2010). Poor pedagogy has also been reported from students across many disciplines in science and mathematics as a primary reason for switching their major (Seymour & Hewitt, 1997). For this reason, calls for novel evidence-based instructional practices in chemistry have resounded from many educational research platforms over the past twenty years (Cooper, 2010; Lloyd & Spencer, 1994; Nameroff & Busch, 2004; National Research

Council, 2003, 2012). Indeed, some have answered this call for innovative teaching approaches in college chemistry, but the evidence put forth regarding their effectiveness is often narrowly focused.

The majority of research on the efficacy of innovative teaching approaches and strategies in college chemistry are either evaluated exclusively based on student performance, or by using questionnaires on student attitudes with little or no evidence of validity or reliability (Dougherty et al., 1995; Farrell, Moog, & Spencer, 1999; Flynn & Biggs, 2011; Hanson & Wolfskill, 2000; Hockings, DeAngelis, & Frey, 2008; Lewis & Lewis, 2005; Paulson, 1999; Rajan & Marcus, 2009; Tien, Roth, & Kampmeier, 2002). Furthermore, most of the performance metrics used for evaluating the efficacy of teaching practices among these studies are course grades and exam grades. This type of evidence is subject to a high degree of variability between different institutions and instructors, and renders the conclusions non-generalizable. A few studies have investigated how certain teaching strategies affect other aspects of students, including their persistence in a two or three semester chemistry track (Geburu, Phelps, & Wulfsberg, 2012; Kampmeier, Varma-Nelson, & Wedegaertner, 2000; Mitchell, Ippolito, & Lewis, 2012), perceptions of the material, and depth of understanding (Hamby Towns & Grant, 1997). However, there is a dearth of research examining the effects of innovative teaching strategies on student motivation in chemistry. Only one study was found that investigated students' motivation, namely, their self-efficacy and interest, from two general chemistry classes with different instructional approaches (Chase, Pakhira, & Stains, 2013). It should be noted,

however, that the researchers in this study used two surveys which together contained 89 items. This would take a significant portion of class time to administer, and in addition, by using lengthy surveys, the rate of incompleteness is likely to rise (Heredia & Lewis, 2012; Lichtenstein et al., 2008). It would be more practical for instructors to have a brief instrument or set of scales that measured several motivational constructs while maintaining adequate psychometric properties. A tool such as this would allow the instructor to collect data multiple times in a single semester without overtaxing the students or using much class time.

Purpose of the Study

Three motivational constructs were chosen for this study and were investigated in the context of general chemistry: self-efficacy (Bandura, 1977), interest (Renninger, 2000), and effort beliefs (Blackwell, 2002; Sorich & Dweck, 1997). These three were chosen based upon their salience in the field of educational psychology, as well as the lack of credible research involving these variables that targets college chemistry students. The overarching theoretical framework that encompasses self-efficacy, interest, and effort beliefs is social cognitive theory. Social cognitive theory posits that individuals' actions are dependent upon a reciprocal causation of personal (cognitive and affective) and environmental factors (Bandura, 1986). In other words, human agency, while internally sourced, is influenced both by internal and external factors. Self-efficacy, as a construct, was spawned from social-cognitive theory and deals with one's belief in his or her ability to accomplish a given task. Also falling under the

broader social cognitive umbrella, effort beliefs represent the extent to which individuals believe their abilities can grow with expenditure of effort. Dweck (2002) explains that individuals either have an incremental (growth with effort) or entity (fixed) theory of intelligence and ability. Finally, the predominant theory on personal interest advances the notion that one's interest is composed of feeling and value-related valences (Schiefele, 1991). In terms of motivation, this means that one is more likely to engage in an activity if he or she enjoys it and/or values it. All of the constructs identified above are well established in the literature and have been the foundation of many studies that have illuminated our understanding of academic motivation.

This study has three main objectives that address the research problem stated above. The first objective was to adapt three scales from existing measures of self-efficacy, interest, and effort beliefs to fit a college chemistry context. This involved minor wording changes to make the items specific to the subject of chemistry. Following the wording changes, the scales had to be pilot tested with the target population - first-semester general chemistry students. The purpose of the pilot test was to establish evidence of validity and reliability. A particular scale cannot be said to measure self-efficacy, for example, among US college students if there is no evidence to support that the scale produces valid and reliable scores (Arjoon, Xu, & Lewis, 2013; Barbera & VandenPlas, 2011). Since none of the scales used in this study have ever been tested with US college chemistry students, each scale was subjected to a rigorous examination of validity and reliability evidence. This evidence was gathered through a mixed

methods approach (Creswell, 2013b) comprised of both quantitative and qualitative data. The quantitative data were gathered in first-semester general chemistry classes by administering the scales. The qualitative strand was conducted using semi-structured interviews with the same population (Creswell, 2013a). Taken together, the quantitative and qualitative approaches provide a more enriched picture of how the scales function in a college chemistry setting than either of them alone.

The second objective of this study was to conduct a classroom-based investigation, whereby students' self-efficacy, interest, and effort beliefs were measured. The main goal for this phase of the study was to evaluate the utility of the modified scales in a quasi-experimental classroom study. The primary purpose behind modifying the scales for chemistry was to provide a tool for practitioners to evaluate various instructional practices based on dimensions of their students' motivation, and not just their performance. Hence, this phase was implemented to examine the functionality of the scales in a setting for which they are intended to be used. Furthermore, the data gathered on motivation, in addition to students' performance in the course, was used to test *a priori* path models. These path models provide quantitative evidence for directional, predictive connections among the motivational variables and course performance.

The third objective of the study was aimed at gaining a deeper understanding of effort beliefs. Although beliefs about effort are conceptualized in the context of implicit theories (Dweck, 2012), which have been extensively

investigated, only three studies have been found that explicitly measure effort beliefs (Blackwell et al., 2007; Jones, Wilkins, Long, & Wang, 2012; Tempelaar, Rienties, Giesbers, & Gijsselaers, 2015). Thus, there is a scarcity of information on how students' effort beliefs toward a specific subject are formed, how they change, and what influences the changes. Although cognitive interviews about the effort beliefs items were performed, another set of interviews were utilized to examine the deeper dimensions of this construct. Students were solicited for interviews based on their effort beliefs score on the scale administered in their general chemistry course. Not only does this research benefit the chemical education community but also the field of educational psychology by providing a rich source of information about a largely understudied facet of motivation.

Research Questions

This study explored the connections among self-efficacy, interest, and effort beliefs with college chemistry students, as well as evaluate the functionality of the scales used to measure these variables in a subsequent classroom-based investigation. This study is framed by the following research questions:

- Q1 What modifications are needed to produce brief, chemistry-specific scales of self-efficacy, interest, and effort beliefs?
- Q2 What evidence supports the functioning of each of the modified scales?
- Q3 To what extent do students' self-efficacy, interest, and effort beliefs change across the first semester of general chemistry?
- Q4 To what extent are students' self-efficacy, interest, and effort beliefs affected by brief interventions targeting their values and implicit theories of intelligence?

- Q5 What are the connections among self-efficacy, interest, and effort beliefs with general chemistry students?
- Q6 To what extent do self-efficacy, interest, and/or effort beliefs predict course performance in general chemistry?
- Q7 What are the sources and influences of effort beliefs toward chemistry among general chemistry students?

Significance of the Study

There is a need in the chemical education community to address the motivational dispositions of chemistry students. Research in other disciplines has consistently shown the importance of motivation in education (Bandura, 1997; Brophy, 2010; Dweck, 1986; Glynn et al., 2009; Singh, Granville, & Dika, 2002; Vallerand et al., 1992). In line with the burgeoning attention given to affective and motivational factors in chemistry, this study offers the chemical education community three major contributions.

First, the results from this study offer instructors of chemistry several new tools for assessing the motivational climate of their classrooms, as well as the motivational impacts of any new teaching approaches. The scales modified and tested herein allow instructors to evaluate students based on their self-efficacy, interest, and effort beliefs in a manner that is neither taxing on the student, nor time-consuming for the instructor. The brevity of these scales also permits instructors to collect data at multiple time points without losing too much class time. In this way, the students' scores can be tracked throughout the semester, shedding light on motivational patterns that could lead to improved teaching approaches. Furthermore, due to the evidence reported in this study for reliability

and validity, users of the scales will gain the assurance that the items have withstood the rigor of several lines of psychometric testing.

Second, this study piloted small changes to a first-semester general chemistry curriculum investigated the effects of these changes on students' performance and motivation. Students in two first-semester general chemistry sections, taught by the same instructor, received two different sets of interventions. The interventions were designed to prompt students to consider their personal values and are intended to be brief with minimal effort needed by the instructor. No study in chemistry has been found that incorporates these interventions, though they have been used in other disciplines with success (Cohen, Garcia, Apfel, & Master, 2006; Miyake et al., 2010). The specific target of these changes was the effect on student performance, as this is how they have been used previously. However, it is plausible that they could also impact students' motivational beliefs, since performance and motivational beliefs are deeply connected (Eccles & Wigfield, 2002; Pajares, 1996; Pintrich & De Groot, 1990).

Lastly, this study provides a path model indicating the direction and strength of the relationships among self-efficacy, interest, effort beliefs, and course performance of chemistry students. No studies have been found in chemistry or any other discipline that incorporate all of the latent constructs to be tested herein. *A priori* path models were based on theoretical relationships among the constructs, where available, as determined from research in other disciplines. In the case of effort beliefs, on which so little research has focused,

the connections were more exploratory in nature. Hence, this study is likely to have implications beyond the field of chemical education. By exploring the connections among latent motivational traits and course performance, more can be understood about how self-beliefs influence cognitive output. This will lead to a greater awareness about specific constructs of motivation that can be targeted to enhance student performance and academic success.

Limitations

This study is subject to the following limitations:

- 1) The samples to be used in this study were convenience samples from two institutions in the same region of the United States. In addition, the students' responses were only used if they grant consent to the researcher. Hence, the responses are unlikely to be representative of general chemistry students everywhere (Crotty, 1998). As a result, the inferences and conclusions drawn from the statistical analyses and interviews might have limited generalizability to the entire population of general chemistry students.
- 2) The interview participants were a small sample of volunteers. As a result, their views and opinions might not be representative of the population of general chemistry students.
- 3) The data reported herein were gathered from general chemistry courses only and are not intended to be generalized to upper-level chemistry courses or to any other subject area.
- 4) The phases of this study where data were collected at multiple time points

over a semester experienced attrition and missing data. Hence, the data set shrunk as the semester progressed due to the responses from missing participants being unobtainable. This has two potential consequences. First, as the sample size is reduced, the statistical power for any tests is lower. This can result in an increase of type II error, which simply means erroneously rejecting a statistically significant result (Cohen, 1992). Second, the sample of students lost might represent an important subset of the population, and their responses might differ from those who regularly attend class.

Definition of Terms

Social cognitive theory – a theory that endorses a model where human agency is determined by a reciprocal causation of cognitive, affective, and environmental factors (Bandura, 1986, 1989)

Expectancy-value theory – a theory of achievement motivation built on the notion that individuals' choice, persistence, and performance in a given activity can be explained by how well they expect to do and how much they value that activity (Wigfield & Eccles, 2000)

Self-efficacy beliefs – judgments of one's capability to accomplish a certain level of performance (Bandura, 1986)

Individual interest – an enduring predisposition for a person to reengage with specific tasks, subject areas, or activities (Hidi, 1990; Schiefele, 1991)

Situational interest – interest triggered spontaneously through an interaction with the environment (Harackiewicz et al., 2008)

Maintained interest – a form of situational interest that endures or is “held” for a period of time and involves focused attention during that time (Hidi & Renninger, 2006)

Implicit theories – “beliefs about the nature of human attributes” (Dweck, 2012)

Entity theory of intelligence – the belief that intelligence is fixed and cannot grow or change (Dweck, 2012)

Incremental theory of intelligence – the belief that intelligence is malleable and can change over time (Dweck, 2012)

Effort beliefs – beliefs about whether or not expenditure of effort will lead to a desired outcome

Latent variable – a variable that cannot be directly observed but must be inferred based on other observable variables

CHAPTER II

REVIEW OF LITERATURE

This review is organized into several sections, and begins with an overview of teaching and learning in chemistry, with particular focus on innovative teaching strategies. Next, a brief history on academic motivation research will be given, highlighting how the field has transformed from being dominated by a few mechanistic theories to a host of cognitive-centered theories. The connection between and importance of motivation and education will be delineated, followed by a discussion of each of the three constructs used in this study (self-efficacy, interest, and effort beliefs) in terms of previous research findings, links with other psychological constructs, and areas for future studies. Published instruments and scales related to the above constructs will be pointed out. Finally, relevant psychometric properties will be discussed, and how inferences drawn from data can be supported by validity and reliability evidence. This review of literature will provide the necessary background to understand the methods employed in this study as well as the research gap intended to be filled by this study.

Teaching and Learning in College Chemistry

Chemistry is a difficult subject for many college students. The cognitive demands range from solving algebraic equations and making sense of abstract chemical representations to comprehending the unfamiliar language of periodic trends and chemical energy. These difficulties are at least partially responsible for the high D,F,W rates of introductory chemistry courses, and low retention of STEM majors. However, the chemistry content may be only one of several obstacles standing in the way of students' success. The traditional method of teaching general chemistry, whereby an instructor teaches a set of historically entrenched topics while students sit mostly as passive observers, has been the source of criticism in the chemical education community (Black & Deci, 2000; Cooper, 2010; Lloyd & Spencer, 1994; Spencer, 1999). Due to the pace of the course, many topics in general chemistry are taught at a superficial level that can, at best, be marginally understood by students before moving on and starting the next topic. This has led to a number of calls to reform both the way general chemistry is taught and the material covered during the course (Cooper, 2010; Council, 2003; Lloyd & Spencer, 1994; Nameroff & Busch, 2004). Although the traditional lecture style remains dominant in most general chemistry classrooms and the content has changed little, some researchers have offered up alternative strategies to teaching the course. These include cooperative learning (Dougherty et al., 1995), peer-led team learning (PLTL)(Gosser & Roth, 1998), process-oriented guided inquiry learning (POGIL)(Farrell et al., 1999), and problem-based learning (PBL). Differing in their implementation and structure, all of these

teaching strategies focus on peer-peer engagement, and student-centered classroom environments.

Cooperative learning has been around since the early 1970s, but was mainly associated with elementary and secondary school grades (Slavin, 1991). The theory behind cooperative learning is that by encouraging students to discuss, debate, and collaborate, they are enhancing their thinking skills and increasing higher order learning (Slavin, 1991). The first published example of cooperative learning in college chemistry was a study by Dougherty and colleagues (1995). The authors of this study argued that there was little substantive quantitative evidence on the efficacy of cooperative learning. This was due to a lack of experimental studies with any type of control group. Hence, the authors utilized three sections at their institution for the study, two of which employed varying degrees of structured cooperative learning activities, and one control section with no cooperative learning activities. Some of these activities included group quizzes, group homework, and intragroup e-mailing of questions/answers related to chemistry. Most of the activities occurred during out-of-class recitation sections, so the lecture format was largely unchanged. It was shown that students in the most structured section were more likely to persist in the class and scored higher on the final exam than students who were in the control section. Other studies have reported positive findings following the implementation of cooperative learning in graduate-level thermodynamics (Hamby Towns & Grant, 1997), organic chemistry (Paulson, 1999), and general chemistry (Geiger, Jones, & Karre, 2008; Kogut, 1997). Cooperative learning

exercises are stand-alone, meaning that the extent to which they are implemented in a class can vary greatly. An instructor may choose to have cooperative learning activities be entirely outside of class, whereas another instructor may choose to completely modify his or her class to incorporate cooperative learning in every period. The following three alternative teaching strategies to be discussed, though called something different, are based on peer-peer cooperative learning.

Process Oriented Guided Inquiry Learning (POGIL) was developed for chemistry by researchers with the belief that students will “construct their own knowledge derived from what they already know” (Spencer, 1999). This belief is the foundation of an epistemological theory called “social constructivism”. Social constructivism is the guiding theory behind much research in science education, with the central principle that students build their own meaning and knowledge based on prior knowledge and their interaction with the environment (Anderson, 2007; Bodner, 1986; National Research Council, 2012). POGIL is intended to be a complete overhaul of the way lectures are traditionally given. Students are split into small self-governing groups, where each student has a specific role. Students spend nearly the entire class period in these groups working on assignments designed to guide them to understanding the chemistry material themselves. The role of the instructor is as a facilitator, who moves among groups guiding discussions and asking questions as a means of clarifying material (Farrell et al., 1999). Although originally developed for small class sizes, POGIL has been adapted for use in large lecture sections as well (Ruder &

Hunnicutt, 2008; Yeziarski et al., 2008). This model has been applied to many areas of chemistry and beyond with mixed success in regard to performance improvements. Farrell & Moog (1999) reported a dramatic reduction in D,F,W rates (21.9% to 9.6%) of general chemistry students following the implementation of POGIL. It has also been shown that students in POGIL classrooms performed better on American Chemical Society (ACS) standardized exams than students in traditional lecture classrooms (Andrei & Emily, 2008; Patrick & Diane, 2008). POGIL has been used in medicinal chemistry with grade improvements reported over previous years where the traditional lecture model was used (Brown, 2010). However, the results from other studies suggest no difference in the course performance of students in POGIL versus traditional classrooms (Chase et al., 2013; Martin & Randall, 2008). As stated earlier, POGIL is intended to replace traditional lectures, and most of the positive impacts reported are from instructors who carried out a massive overhaul of their teaching practice. This can be a difficult adjustment for many college instructors and departments to take on due to the drastic changes and learning curve of this teaching method. Some instructors may choose to only include POGIL partially in their course. For example, Chase et al. (2013) incorporated POGIL only in their recitation sections of general and organic chemistry, and they reported no grade difference between students in POGIL and traditional sections of recitation. In another study, researchers used POGIL only in organic chemistry lab and reported higher scores, from previous years, on a focused set of questions related to mechanisms (Schroeder & Greenbowe, 2008). Although the results on

performance improvements are mixed, it is worth noting that the methods of implementation are not homogeneous across the studies. Despite this, POGIL remains one of the most well known alternative teaching strategies in chemical education.

Problem-based learning (PBL) was first developed and used with medical students (Barrows & Tamblyn, 1980; Schmidt, 1983). PBL arose out of the concern that students in medicine or other disciplines, such as physics, were unable to apply the knowledge they possessed in a real-world context (Schmidt, 1983). PBL has also been described as a tool for learning concepts and skills that has not been taught previously (Dods, 1996). As with POGIL, PBL is based on a constructivist theory of learning (Savery & Duffy, 1995). PBL usually involves breaking students into small groups and giving them an authentic problem they might encounter in the context of the subject matter. Typically, all students in the class are given the same problem to solve, but one application of PBL in chemistry has introduced variation among different groups (Overton & Randles, 2015). Problems given to students are as diverse as the subject areas from which they originate. For example, a problem given to medical students might be a scenario where a patient comes into a hospital complaining of a series of symptoms. The students are then charged to come up with an explanation of the symptoms, then provide recommendations for physical examinations, laboratory tests, and treatments (Schmidt, 1983). In a biochemistry lecture course, the problem could resemble a real-world scenario whereby students are role-playing as researchers and given the task of introducing a

metal-binding site to a protein. The students must know which amino acids will bind metal centers, choose the most appropriate location for this amino acid, then analyze the problems associated with predicting the conformation of the protein at a given site (Dods, 1996). With reference to published literature, PBL is a much more popular teaching approach in laboratory settings than in lecture (Flynn & Biggs, 2011; Hicks & Bevsek, 2011; Kelly & Finlayson, 2007, 2009; McDonnell, O'Connor, & Seery, 2007; Nielsen, Scaffidi, & Yeziarski, 2014; Ram, 1999). This is perhaps due to the fact that authentic scenarios are better manifested and experienced in the laboratory with equipment and instruments than in lecture with pen and paper. The impacts of PBL on student performance are less reported than with either POGIL or cooperative learning. Most studies either did not measure differences in performance (Dods, 1996; Hicks & Bevsek, 2011; Kelly & Finlayson, 2007, 2009; McDonnell et al., 2007; Overton & Randles, 2015; Ram, 1999), or suggested that students' grades were comparable to previous semesters (Flynn & Biggs, 2011). This is likely due to the low rate of PBL implementation in lecture. However, many studies did find that students enjoyed PBL or had positive things to say about the format (Flynn & Biggs, 2011; Hicks & Bevsek, 2011; Kelly & Finlayson, 2009; McDonnell et al., 2007; Overton & Randles, 2015; Ram, 1999).

Another alternative teaching approach that has become popular in college chemistry courses is peer-led team learning (PLTL). Just like POGIL, PLTL was developed specifically for college chemistry (Woodward, Weiner, & Gosser, 1993), although it has been used in other subjects such as biology, mathematics,

and anatomy (Born, Revelle, & Pinto, 2002; Hughes, 2011; Quitadamo, Brahler, & Crouch, 2009). Within the discipline of chemistry, the implementation of PLTL centers on general (Gosser & Roth, 1998; Hockings et al., 2008; Lewis, 2011; Quitadamo et al., 2009) and introductory organic chemistry courses (Kampmeier et al., 2000; Wamser, 2006). Peer-led team learning involves training peer leaders, usually undergraduates who have done well in the course previously, to lead weekly discussion sessions with groups of 6-8 students (Gosser et al., 1996). PLTL is not designed to replace lecture completely, but rather to augment it by exchanging one hour of lecture for one or two hours of team learning outside of the lecture period. Problem sets for the team learning are usually given by the instructor, and peer leaders typically work closely with instructors on learning goals. One aspect that may pose a hurdle for instructors and departments is the cost associated with employing peer leaders. Due to the time commitment associated with attending the team learning sessions, the peer leaders are often paid a stipend for the semester. For a section of 96 students, 12 peer leaders would be needed if PLTL were to be properly implemented at the maximum student to leader ratio of 8:1. The total cost to employ the peer leaders for the semester, assuming \$500 per leader, is \$6000 (Quitadamo et al., 2009). Furthermore, at many schools, the enrollment for general chemistry is much higher than 96 students. The cost alone may exclude many departments from implementing PLTL. With that said, there is another way to incentivize the workload of peer leaders. Some departments have offered course credit to the peer leaders in exchange for leading a group during the semester (Mitchell et al.,

2012; Quitadamo et al., 2009; Tien et al., 2002). In addition, to reduce the number of peer leaders needed, higher ratios of peer leader to students have been used (Lewis, 2011; Mitchell et al., 2012). In spite of these alterations and adaptations, PLTL is still a large undertaking as an instructor. Recruiting and training peer leaders prior to and during the semester is a substantial time commitment. Some researchers and instructors believe the time commitment to be justified because of the gains in student learning and performance. Many published studies have reported positive performance impacts of PLTL among students in general and organic chemistry (Hockings et al., 2008; Lewis, 2011; Mitchell et al., 2012; Tien et al., 2002; Wamser, 2006). In addition, students generally responded positively toward PLTL (Hockings et al., 2008; Tien et al., 2002; Wamser, 2006).

The aforementioned teaching strategies represent the most structured and publicized alternatives and additions to traditional lecture style instruction in college chemistry. Other strategies that have been used in college chemistry include the use of clickers (MacArthur & Jones, 2008) and flipped classrooms (Herreid & Schiller, 2013). Even with the abundance of evidence in the literature for these practices, the majority of instructors still rely on the traditional model of lecturing, note taking, and memorization. The barriers associated with shifting the current “lecture-only” paradigm are not very well documented among chemistry faculty. However, research from other science fields can shed some light on why so few science faculty members adopt evidence-based instructional strategies. For example, in physics education, researchers have found that the most

frequently cited barrier as to why instructors avoid implementation of research-based teaching strategies is the time commitment (Dancy & Henderson, 2010). It takes time to learn about the strategy, time to prepare materials for class, and time during class. Many faculty members might be interested in innovative teaching practices, but are unable or unwilling to commit the necessary amount of time to implement them due to other constraints such as research, mentoring students, or high teaching loads (Fairweather, 2008). Other barriers include situational factors such as room layout, class size, departmental norms, and expectations of content coverage (Henderson & Dancy, 2011). However, if the instructor is not convinced the teaching strategy is effective, the barriers mentioned here are never even encountered.

The evidence used by most reformers of college chemistry instruction to support the efficacy of a specific practice is overwhelmingly performance-based and subjective in nature. For example, the most cited positive impact of innovative teaching strategies is “retention” or lower D,F,W rates. “Retention” is synonymous with “success rate” in the chemical education literature, usually referring to a comparison of students who pass the course with a C or better with those who earn a D,F, or W in the course (Dougherty et al., 1995; Lewis, 2011; Mitchell et al., 2012; Tien et al., 2002). Some might argue it would be simple to tweak the content or grading so as to inflate the impact of an innovative teaching strategy. Indeed, some studies have used weak evidence of performance gains by comparing total scores or D,F,W rates after implementation of a teaching strategy to those before implementation (Farrell et al., 1999; Tien et al., 2002).

By putting forth this evidence, the reader is left to guess whether all measures and expectations for the course remained constant after the new teaching strategy was implemented. To ameliorate this concern, some authors use equivalent exams between treatment (alternative teaching strategy) and control (traditional teaching method) groups (Lewis & Lewis, 2005), or the standardized ACS exam scores as test variables (Andrei & Emily, 2008; Lewis, 2011; Lewis & Lewis, 2005; Mitchell et al., 2012; Patrick & Diane, 2008; Wamser, 2006). By using equivalent measures for both the treatment and control groups, the performance results have added credibility. Nevertheless, the evidence still might not be convincing for some instructors. For example, a statistically significant difference in ACS exam percentile rankings could equate to one additional question, on average, being answered correct by the treatment group. Even if the alternative teaching strategy is the cause for this difference, instructors may not find it worth their time to implement this strategy for such a slight gain in student performance. Most of the performance impacts cited in the literature are marginal at best, and associated with small sample sizes. Clearly, the evidence put forth thus far is not compelling enough to cause a paradigm shift in the way chemistry is taught at most colleges and universities. Instructors are not going to put forth the time and effort to overhaul their teaching practice for a few extra points on a final exam. Evidence of improving more persistent qualities and dispositions within the student, such as confidence or interest, could broaden the dimensions of efficacy for a given teaching strategy.

Most researchers and instructors who have published studies on alternative teaching strategies in chemistry neglect to study the impact that their approach to instruction has on students' interest, motivation, and attitudes toward chemistry. However, a few notable examples do exist. Chase et al. (2013) used two attitude measures and an expectancy measure to determine if students scored differently based on whether they were in a POGIL course or traditional lecture. Although research-based, these measures were lengthy, adding up to 89 items between the two. This would take up a substantial amount of class time and likely be exhausting for students. It has been found that completion and participation rates tend to be lower with more time consuming surveys (Heredia & Lewis, 2012; Lichtenstein et al., 2008). Other researchers have either developed "in-house" attitude and motivation measures, but put forth no evidence of validity or reliability for the scores obtained (Hockings et al., 2008; Overton & Randles, 2015; Tien et al., 2002), or used an instrument inappropriate for the target population (Dougherty et al., 1995). Fairweather (2008) points out,

The usefulness of any assessment technique ultimately depends on both its rigor *and* ease of use...Researchers and faculty members who customize evaluation tools for idiosyncratic applications are not likely to find an enthusiastic response from colleagues.

To produce credible results for any research study, the tools used to measure the variable(s) of interest must be high quality, and subjected to an evaluation prior to their use. Failure to do so renders the test results, and the inferences that follow, questionable at best (Blalock et al., 2008).

Although there have been efforts by chemical education researchers to develop or modify quality measurement tools for motivational and affective

constructs, the majority of these are lengthy (Barbera et al., 2008; Bauer, 2005, 2008; Dalgety et al., 2003; Uzuntiryaki & Aydin, 2009). This requires instructors to give up valuable class time, and students to remain attentive and thoughtful in their responses for an extended period. Hence, it would be useful for instructors to have brief, yet complete, measures available that target specific motivational and affective constructs in chemistry. There are a few examples of researchers in chemical education using short and concise scales to gauge these types of constructs. Xu & Lewis (2011) reconstructed a chemistry attitudes measure published by Bauer (2005) by shortening it from 20 items across five subscales to eight items across two subscales. They were able to show through a rigorous psychometric evaluation of validity and reliability that the original 20-item measure was redundant and the scales were misaligned. This work provided a notable example of what could be done to make an instrument easier to administer and the results more interpretable for instructors of chemistry. In another study, Villafañe et al. (2014) used a truncated version of a self-efficacy scale from an instrument developed for college chemistry by Dalgety et al. (2003). The original self-efficacy scale contained 12 items, but the authors chose five items out of the scale that best matched what their participants would encounter in chemistry class. In doing so, they were more easily able to administer the shortened scale at multiple time points throughout the semester. As with the Xu & Lewis (2011) study, this study also provided structural validity evidence that the truncated scale functioned adequately with the target population. These studies unfortunately represent the majority of the work that

has been done in chemical education to produce brief scales that can be easily and quickly administered in the classroom to gauge the motivational and affective climate. There are many other variables related to motivation that could be measured, and could provide instructors with valuable feedback about how their teaching methods influence students' beliefs. Motivation is a key component to the success of any student, and having quality assessment tools to measure it will help steer instructors of chemistry toward strategies that promote and foster a positive motivational climate.

A Brief History of Academic Motivation Research

During the mid-twentieth century, motivation research held a dominant place in the field of psychology. The word “motive” has Latin roots, and means, “to move”. Hence, psychologists, at the time, were concerned with the processes necessary to cause stationary organisms to engage in activity. Early theories of human and inhuman motivation were based upon conceptions of homeostasis and drives. Drives, the psychological equivalent of needs, are the necessities of life such as hunger, thirst, and sex. Homeostasis was conceptualized as a state where equilibrium was achieved because of an action on the part of the organism to reduce a particular drive. For example, a wolf in need of food attains and consumes the food to reduce the drive of hunger and returns to homeostasis.

In the 1941 and 1950 editions of the *Encyclopedia of Educational Research*, the cited research centering on motivation tended to utilize animal models as a means for explaining human behaviors. The reason for this was the availability of animals for experimental research and the liberty that could be

taken with animals during these experiments. Examples of cited experiments included depriving food from rats and electrically shocking mice and other animals (Weiner, 1990b). By depriving food from rats, researchers were interested in what the activity level, or motivation level, of the rats would be after not eating for several days. The electric shock experiment was aimed at understanding the relationships of punishment and incentives with the speed of learning. Would the rats learn faster if they were punished with an electric shock? What degree of electric shock and how much incentive produce the most optimal gains in learning? The experiments with animal models were not conducted to understand the motivation of rats and mice, but for a more general purpose. As pointed out by Young (1941), "The work with animal subjects provides a biological perspective for studies of human beings, and often it reveals clearly the fundamental and general principles of human motivation" (p.736). Theories were built based on the results of these experiments and generalized to humans, whose behaviors were considered too complex to study at the time (Weiner, 1990a).

During the 1930's, for example, it was suggested that learning can occur without an increase in motivation. Motivation was conceptualized as a response to drives (hunger, thirst, defense, etc.). Without a drive reduction, or incentive, a change in behavior would not occur. However, research by Tolman (1932) indicated that new learning was separate from motivation, after experiments with animals showed that they could learn the structure of a maze and develop habits without incentives. Incentives did, however, increase the performance of animals

in the maze. Thus, it was concluded that new learning does not depend on motivation or drive reduction, but performance increases with motivation and rewards (Marx, 1958; Weiner, 1969, 1990a). This was a problem for the educational psychologists at the time because learning was a key indicator of one's level of motivation (Weiner, 1990a). Even though this theory was based on animal models, the implications spilled over into human motivation research and complicated the work of psychologists interested in understanding how to promote new learning among students. This is just one example of the overlap that occurred between animal-based motivation research and human motivation research.

Although research based on animal models made up the bulk of what was published on motivation during the mid-twentieth century, research aimed at academic motivation was also reported. Some of the experimental manipulations in educational research that occurred at that time would, today, be considered unethical and even potentially harmful to children. This is no surprise, as the field of psychology during that era was steeped in primitive and subhuman conceptions of behavior and the forces that drive behavior. Nevertheless, there are five major topics that encompass the educational research reported by Young (1941, 1950) in the *Encyclopedia of Educational Research*. These include: praise and reproof, success and failure, knowledge of results, cooperation and competition, and reward and punishment. A few examples of the major educational research on motivation reported in the 1941 and 1950 volumes of the *Encyclopedia of Educational Research* are highlighted below.

An example of a major study on praise and reproof of children was conducted by Hurlock (1925). The study included 106 children from grades 4 and 6 and examined the success rate and improvement over time on an addition test of children who were consistently praised, reproofed, or ignored. In addition, a control group was formed with students who received no special or unusual treatment and were separated from the rest of the students. Hurlock (1925) found that students who were consistently praised in front of the class made greater gains than those who were reproofed in front of the class or ignored altogether. However, the control group performed the worst among the four groups, behind the ignored group. Thus, it was concluded that both praise and reproof are effective incentives for higher achievement, but praise is superior to reproof.

Sears (1937) conducted a study of 19 college freshmen on whether reported improvements or decrements in speed with a card-sorting task resulted in actual increases or decreases in speed. Participants were divided into two groups: success ($n = 9$) and failure ($n = 10$). Participants were asked to sort a deck of 52 cards into the four suits and, prior to each trial, set a goal for the time it would take them to complete the task. In order to generate feelings of success and failure among the participants, the times were falsely reported in both groups. For the success group, the individual times were reported to be faster than the goals set by the participants. For the failure group, the opposite was true. The actual results indicated that those in the success group improved overall throughout the day in successive trials and for the three days the experiment took place. On the contrary, those in the failure group declined in

performance by taking more time to sort the deck in successive trials throughout each day. However, over the three days, the average sorting time did improve for the failure group. Although not directly linked to the educational context, the connection could be made that knowledge of success and failure, whether true or fabricated, can contribute to one's actual performance in school. Later theories in educational psychology would build upon the notion that past successes and failures shape one's beliefs in their abilities and influence their performance and persistence in the face of failure.

By the mid-1950s, the focus on motivation research was largely unchanged from previous years. "Motives" were defined and characterized in the context of the "presentation of a particular stimulating situation to an organism and the observation of characteristic approach or withdrawal behavior" (Marx, 1958, p. 889). The majority of the research centered on experimental manipulations involving animals or young children, and mechanistic explanations of their behavioral responses. However, some theories associated with cognitivism, or choice of behaviors, were just starting to gain ground. For example, Tolman (1955) formulated a theory that described behavior as purposive and a function of organismic demands, the expectancy that the response will lead to the goal, and the value of the goal. Hence, behavior is dependent not only on drives or needs, but is also directed by the expectancy of success and value of the goal. Another similar theory, posited by Atkinson (1957), explains behavior in terms of a multiplicative function including motives, expectancy, and incentives. The motives are either the motive to achieve or the

motive to avoid failure. Atkinson (1957) argued that those with a strong motive to achieve prefer tasks of intermediate risk, and those with a strong motive to avoid failure prefer very easy or very difficult tasks. Both of these theories would later become known as expectancy-value theories of motivation.

In terms of educationally relevant motivation research, there was little growth during the decades between 1940 and 1960. Marx (1958) expresses his concern by stating, “we are handicapped by the relatively small amount of relevant material made available in the educational literature” (p. 895). He goes on by highlighting the “discrepancy between the widespread recognition of the motivational problem in education...and the inadequate experimental attention accorded it...”(p. 895). Despite the dearth of relevant motivation-related research in education, some important ideas had emerged that would help shape future theories in educational psychology.

The first had to do with the use of rewards or incentives in the classroom. Researchers had become aware at the time that external rewards (e.g., prizes) should be used carefully to prevent the negative motivational effects from exclusivity or indiscriminate use. Intangible rewarding, such as praise and the facilitation of regular experiences of success, were also regarded as important for student learning. Second, continuous performance feedback was mentioned as a “very important motivator in its own right” (Marx, 1958, p. 896). Students who are aware of their successes and failures continually are able to make appropriate adjustments in their performance to place themselves on a path to success. Lastly, what would now be called inquiry learning was proposed as a way to

motivate students by holding their attention in suspense until they were able to make a learning discovery themselves. These three notions about student learning and motivation are evidence that new theories and a deeper understanding of student motivation were on the horizon.

During the decades that followed Marx's (1958) chapter on motivation in the *Encyclopedia of Educational Research*, research on motivation shifted dramatically from a mechanistic view to a cognitivist view of motivation. A major theory that emerged in the 1960s was Atkinson's (1964) conception of achievement behavior as stemming from the approach-avoidance conflict. According to this view, individuals either display a motive to approach success or a motive to avoid failure when confronted with a task. If their motive to approach success is greater than their motive to avoid failure, then individuals are likely to approach achievement-related tasks. The implications of this theory were far-reaching and helped lay the foundation for future achievement theories that focused on individual choices, responses, and preferences toward various educational demands. Also during the late 1950s and 1960s, researchers made use of existing instruments and developed new instruments to measure constructs associated with achievement strivings, anxiety, and avoidance behavior. This trend would continue and grow to become the major mode of measuring variables linked to motivation.

In the 1970s and 1980s, the earliest theories of mechanistic motivation that focused on drives, needs, and arousal had all but faded away in the shadow of a new paradigm of conceptualizing human motivation which emphasized self-

beliefs, causes of failure and success, and personal responsibility (Weiner, 1990b). As Weiner (1990b) puts it,

In sum, for motivational psychologists there have been fundamental shifts in theory and research focus over the past few decades, and the basic metaphor for what it means to be human has shifted from robotic machine to a scientist and/or decision-making economist.

The major contributions to motivation research that were conceived and/or developed during 1970s and 1980s were: self-efficacy theory as part of the larger social-cognitive theory (Bandura, 1977, 1986), Nicholls' (1984) conceptions of ability and task choice to explain achievement motivation, attribution theory (Weiner, 1985), learned helplessness (Seligman, 1975), theories of intelligence and goal theory (Dweck, 1986), and expectancy-value theory (Eccles et al., 1983). All of these theories and ideas, among others developed later (Ryan & Deci, 2000a; Schiefele, 1991), have weathered the passage of time and are now considered the foundations of modern achievement motivation research (Brophy, 2010; Wigfield & Eccles, 2002b). Due to the vast array of theories and conceptions that exist to explain achievement motivation, only those relating to the work of this dissertation study will be presented in detail.

The focus of this dissertation relates to three major constructs of motivation: implicit theories of intelligence (Dweck, 2012), self-efficacy (Bandura, 1997), and interest (Schiefele, 1991). These constructs were chosen out of many others linked to motivation due to their salience in the literature, potential predictive power in chemistry classrooms, and distinctiveness from each other. Although work with self-efficacy and interest has been done in the domain of college chemistry before, the strength and validity of some of these studies is

questionable as it relates to measurement of the constructs. Moreover, no research has been found relating to implicit theories of intelligence or effort beliefs in chemistry, even as this work has gained momentum recently in other domains of education and beyond (Burnette, O'Boyle, VanEpps, Pollack, & Finkel, 2013; Chen & Wong, 2014; Komarraju & Nadler, 2013; Rattan, Good, & Dweck, 2012; Sevincer, Kluge, & Oettingen, 2014).

Motivation research is exceedingly multi-faceted and there exists overlap among many of the prevailing theories. Efforts have been made to introduce parsimony into achievement motivation research. For example, Eccles & Wigfield, (2002) attempted to explain how a plethora of motivational constructs fit into the expectancy-value framework. However, loss of meaning and diffusion are unavoidable when the theoretical framework is too broad. By collapsing so many constructs and motivational processes into one theory, the interpretation and understanding of each element becomes clouded. Thus, in order to strengthen clarity and minimize redundancy, each construct related to this dissertation will be defined and described with linkage to the research and theories from which it has been most comprehensively elucidated.

Self-Efficacy

Overview of Social Cognitive Theory

Social cognitive theory is rooted in the idea that humans are neither autonomous agents, nor mechanical agents of their own behavior when responding to external stimuli. As fully autonomous agents, human behavior would only be subject to an “autonomous inner man”, and void of influence by the environment

(Bandura, 1986). On the other end of the spectrum, mechanical agency proponents argued that the thoughts and emotions of humans are bypassed, as behavior is only a response to external stimuli with future behavior being governed by reinforcement (Bandura, 1986). The mechanical model of agency assumes unidirectional influence in determining human behavior that is an automatic response to a stimulus. The stimulus determines the behavior, and the consequences of the behavior shape and control future behavior. Thus, the stimulus manipulates behavior without the input of thoughts or inner feelings, which were thought of as passive byproducts of the stimulus (Bandura, 1986). Radical behaviorism, along with its tenets of mechanical agency, have faded from mainstream psychology and given way to a more holistic view of behavior determinism.

Social cognitive theory posits an interactive agency, and explains human motivation, thought, and action through a model of triadic reciprocal causation involving environmental events, personal factors, and behavior (Bandura, 1986, 1989). The reciprocal causation, or determinism, assumes a bidirectional flow and interaction between behavior and its sources of influence. Thus, the processes that determine one's volition and consequential behavior are both internally and externally sourced. Humans are not merely passive observers of the environment and recipients of stimulation; but instead, humans both shape and are shaped by their environment through the sensory, motor, and cerebral systems (Bandura, 2001). As Pajares (1996) points out, "individuals are viewed as both products and producers of their own environments and their social

systems” (p. 544). Thus, there exists both an outward flow of action from the mind to the environment, via the motor system, as well as an inward flow of information from the senses to the mind from the environment. Social cognitive theory brought a new perspective on human behavior, one that is governed by the dynamic interaction of thought, action, and environment.

Central to social cognitive theory is the perspective that humans can be agents of their own behavior. Agents are those with the capability to act intentionally. A physical reflex is not an intentional act, and hence, the person experiencing the reflex is not an agent of that action. However, a student who chooses to review for an exam is an agent of his or her studious behavior. Although many animals have the power to act with intent, the following discussion will be framed in terms of human agency.

The core features of human agency are intentionality, forethought, self-reactiveness, and self-reflectiveness (Bandura, 2001). To be an agent of one’s behavior requires intent, which centers on a commitment to see the action come to pass. Intention is a mental representation of future behavior and requires a plan of action. Forethought includes goal setting, identifying potential consequences of the behavior, and setting courses of action to maximize beneficial effects and minimize detrimental effects (Bandura, 2001). As part of forethought, individuals construct cognitive depictions and expectations of future outcomes of their behavior. Forethought, then, can become a powerful motivator to either approach or avoid certain behavior. For example, if a student desiring a passing grade in a class believes that completing a project will lead to a passing

grade, he or she is more likely to finish the project. However, by simply constructing a positive outcome expectation and generating a plan of action will not lead to the passing grade. The agent must also implement the act in a way consistent with the expectation by self-regulating his or her behavior. This link between the outcome expectation and the action is known as self-reactiveness. Self-reactiveness involves the ability to self-regulate motivation and the execution of a given task. The processes associated with self-reactiveness are self-monitoring, adherence to personal standards or goals, and corrective self-reactions (Bandura, 1986). Monitoring one's actions and aligning with predetermined goals provide direction and meaning to one's actions. In addition, goal setting can provide inducements to continue with effortful action until the goal is attained (Bandura, 1977). Corrective self-reactions are the mechanism by which one is able to stay on course in achieving their goals. Where self-reactive processes tend to occur during an action or task, self-reflectiveness will occur outside of the act itself, either before or afterwards.

Self-reflectiveness is a faculty unique among humans whereby they are able to assess their motivations, actions, and values. According to Bandura (2001),

In this metacognitive activity, people judge the correctness of their predictive and operative thinking against the outcomes of their actions, the effects that other people's actions produce, what others believe, deductions from established knowledge and what necessarily follows from it (p. 10).

What people believe about their capabilities, the consequences of their actions, and how these actions fit into the social fabric of their environment are incredibly

influential on a person's behavior. Individuals must rely on personal judgments of themselves and their environment to direct and self-regulate their future behavior. It is these judgments that form self-beliefs which give people the confidence that they possess the ability to control their own conduct and to some extent, their environment. According to Bandura (1997), "People's level of motivation, affective states, and actions are based more on what they believe than on what is objectively true" (p. 2). The specific self-referent beliefs that deal with personal judgments of one's capability to organize and execute a given task are termed self-efficacy beliefs.

Self-Efficacy Beliefs

As a construct, self-efficacy beliefs were first formally defined by Albert Bandura in 1986 as "people's judgments of their capabilities to organize and execute courses of action required to attain designated types of performances" (Bandura, 1986, p. 391). Self-efficacy lies at the core of social cognitive theory, because it is the basis of human agency (Bandura, 2001). Bandura (1986) argues that self-efficacy, as part of a complex system of self-referent beliefs, mediates the relationship between thought and action. Efficacy beliefs influence what tasks a person will undertake, effort expenditure, persistence in the face of aversive experiences, and the types of skills one acquires (Bandura, 1977; Bouffard-Bouchard, 1990; Schunk, 2008). Efficacy beliefs are task-specific, and should not be measured in general terms, as this would obscure what is being assessed by allowing for unnecessary ambiguity (Bandura, 1986; Pajares, 1996).

People would tend to gauge their self-efficacy without a clear task or goal in mind, which could call into question the meaning behind such an assessment.

Bandura (1986) states, "Among the different aspects of self-knowledge, perhaps none is more influential in people's everyday lives than conceptions of their personal efficacy" (p.390). In education, self-efficacy is particularly important, as it has been associated with motivation, in terms of students' effort (Collins, 1992 as reported in Pajares, 1996), persistence (Lent, Brown, & Larkin, 1984), emotional reactions to tasks given in school (Pajares, 1996), college major choice (Hackett & Betz, 1989), and academic performance (Lightsey, 1999; Zimmerman, Bandura, & Martinez-Pons, 1992). As the effects and impacts of self-efficacy have important implications for educational research, so do the sources of self-efficacy.

Sources of Self-Efficacy

There are four sources of self-efficacy theorized by Bandura (1997): enactive mastery experiences, vicarious experiences, verbal persuasion, and physiological and affective states. Not only do these four experiences represent the sources of self-efficacy expectations, but they represent the vehicles for changing one's self-efficacy as well (Bandura, 1977). Bandura's hypothesized four sources of self-efficacy have been confirmed in many academic studies since his initial work (Usher & Pajares, 2008). The four sources of self-efficacy are intertwined and organized within the individual with differing weight placed on each depending on the situation. In some cases, physiological and affective states may play a larger role in determining one's self-efficacy than mastery

experiences, or vice versa. However, in all cases, the individual must process the information from these four sources through reflective thought before they can be used to influence behavior (Bandura, 1997).

Enactive mastery experiences are considered to be the most influential source of efficacy information (Bandura, 1997). This is due to the powerful influence of successes and failures on one's belief in their own capabilities. Success tends to bolster one's confidence and efficacy beliefs toward a task, and failures tend to undermine them. A college chemistry student who has succeeded in high school chemistry will tend to have a higher chemistry self-efficacy than another college chemistry student who was less successful in high school. However, successes and failures are not direct causes of changing one's self-efficacy; rather, they are contributors. More important than the successes and failures themselves, is how they are weighted and interpreted by the individual (Bandura, 1982). Mastery experiences represent a vital personal achievement component of self-efficacy development, and they can be enduring influences. More temporary personal factors are the physiological and affective states of individuals.

A chemistry student who has achieved high marks in her class all semester, but has been sick for a week in advance of her final exam may report a lower chemistry self-efficacy than if she were healthy. Moreover, her sickness might contribute to increased anxiety and stress, which could further lower her confidence to perform well on the exam. The physiological and affective states of an individual are certainly more variable and dynamic factors than the other three

contributors to self-efficacy. For this reason, it cannot be altered externally very easily, and is dependent on myriad of factors independent of the task toward which the self-efficacy beliefs are directed. However complex affective and physiological states are, they cannot be ignored as sources of self-efficacy, as they play an important situational role in executing a given task.

The former two influencing components represent internally sourced factors produced by an individual. The remaining sources of self-efficacy, vicarious experience and verbal persuasion are externally sourced. Vicarious experience refers to modeled attainments by others. The most influential vicarious experiences are those in which the model is perceived to be of similar ability to the observer. For example, peer instruction or group work in a chemistry class could enhance one's self-efficacy through vicarious experience more so than if the instruction came from the professor. By observing successful attainments by peers, students convince themselves that they are capable, as well, to enhance their own performance (Bandura, 1982; Schunk, Hanson, & Cox, 1987). Another facet of vicarious experience is norm-referenced perceptions of success. People tend to compare their own performance with that of others to judge whether they have succeeded or failed. For example, a student who scores a 102 on an exam has no idea if this is a good or bad score until it is compared with the scores of others. This can have a profound impact on one's perceived efficacy, by leading to conceptions of personal deficiencies, in the case of failure, or capabilities, in the case of success. Modeling by others is not simply a means by which to appraise one's ability in the context of a social structure, but

it is instructive as well. By observing others perform certain tasks, the observer is better able to emulate the model both through enhanced confidence and knowledge acquisition.

Verbal persuasion is when a significant person in the life of an individual provides verbal affirmation that the individual is capable of being successful at a given task. Verbal persuasion can have a significant impact on one's self-efficacy by encouraging the person to enact behaviors to counter the difficulty. Bandura (1997) states, "People who are persuaded verbally that they possess the capabilities to master given tasks are likely to mobilize greater effort and sustain it than if they harbor self-doubts and dwell on personal deficiencies when difficulties arise" (p. 101). This effect is enhanced if the persuader has experience in the domain of interest, because experience will confer a degree of credibility to the persuader. Thus, a professor's persuasion may have more of an impact than a fellow student's persuasion because presumably the professor has seen many successes and failures in his class, is acutely aware of what is required for success, and is a better diagnostician than the student.

The integration of these four sources of self-efficacy forms a complex fabric of information organized by cognitive processing that differs by domain and situation. In some instances, more weight might be placed on mastery experiences than verbal persuasion to produce self-appraisals. In other cases where mastery has never been achieved, the person may rely on vicarious experiences and verbal persuasion to formulate beliefs about his or her capability to execute the task. In all cases though, self-efficacy beliefs are a product of

complicated cognitive processing that judges, weighs, and integrates information from a diverse range of sources.

Constructs Related to Self-Efficacy

Several psychological constructs tied to motivation are closely related to self-efficacy; namely, outcome expectations, self-concept, perceived control, and self-esteem. The following discussion will be constrained to outcome expectations and self-concept, as these constructs are most relevant to academic outcomes. The purpose of this section is to shed light on these select constructs in the motivation literature which share the most overlap in meaning with self-efficacy. Due to the relative prevalence of research on self-concept compared to outcome expectations when examining academic outcomes (Bong & Clark, 1999), the most attention will be given to differences between self-efficacy and self-concept in this discussion.

Prior to Bandura's (1977) conceptualization of self-efficacy, he worked toward measuring outcome expectations for fear-based tasks from people who had a particular phobia. Outcome expectations are beliefs about the outcome that a specified behavior will produce (Bandura, 1977), whereas self-efficacy deals with one's judgment of their capability to complete a task. An example of an outcome expectation might be "By studying hard, I will be successful in this class." A related, but distinct self-efficacy item might read, "I am confident I will be successful in this class." While studying phobias, Bandura found that by only measuring what outcome is expected when a certain behavior is put into action

predicts very little about that person's confidence to enact said behavior (Zimmerman, 2000). Hence, this led Bandura (1986) to contend that self-efficacy is a stronger predictor of performance than outcome expectations, a notion that has been empirically corroborated by others (Greene, 1985; Shell, Murphy, & Bruning, 1989).

Self-concept is another related construct to self-efficacy, and has been the subject of much research across many academic domains, including mathematics (Pajares & Miller, 1994), chemistry (Bauer, 2005), english (Marsh & Yeung, 1998), and general science and writing (Pajares, Britner, & Valiante, 2000). Self-concept refers to a general perception of oneself that varies based upon domain (Huang, 2011). In the context of education, self-concept has been described as a generalized "academic self-efficacy" (Pajares & Miller, 1994). For example, someone with a high math self-concept might make the statement "I am good at math." Self-concept, like self-efficacy has been shown to be predictive of performance, but as Bandura (1997) puts it, "self-concept loses most, if not all, of its predictiveness when the influence of perceived efficacy is factored out" (p.11). As such, there is some redundancy in the meaning between the two constructs, which could be easily confused as being the same. However, a careful examination of the nature of these two constructs reveals key differences.

The first notable difference between self-efficacy and self-concept has to do with the clarity in which each construct has been defined. Historically, operational definitions of self-concept have been plagued by inconsistent and

ambiguous terms. Early researchers, Shevelson, Hubner, and Stanton (1976) defined self-concept as “a person’s perception of himself.” They go on to say that self-concept is influenced by interactions with the environment and “significant others”. Wylie (1979) defined self-concept as “cognitions and evaluations regarding specific aspects of the self.” These vague definitions have led to a myriad of interpretations and scales that differ in what they are actually measuring. Bong and Clark (1999) point out that one reason for this could be that there are so many instruments that purport to measure academic self-concept. Many of them include items that prompt students to compare themselves with their peers, and focus on emotions and feelings toward their academics. This differs from how self-concept is viewed currently, with more emphasis on competence judgments. Furthermore, the theoretical construct underlying self-concept has been given a number of different names. In a meta-analysis of 128 studies of academic self-concept, Hansford and Hattie (1982) identified 15 different terms for self-concept. Despite the unclear theoretical representations of self-concept in the past, Bong and Clark (1999) suggest that recent trends in self-concept research have shifted away from feelings and emotions about academia and more toward beliefs about ability or competence in academic domains.

A second theoretical difference between self-concept and self-efficacy is the task-specific nature of self-efficacy versus the broad, domain-related nature of self-concept. An example question for a self-concept measurement item could be “Are you good at chemistry?” This item represents one’s beliefs about their competence in chemistry as an entire domain. To measure self-efficacy in

chemistry, questions would need to be worded much more specifically. For example, “To what extent can you explain laws and theories related to chemistry?” Self-efficacy can only be measured appropriately if it is tied to a specific task, and conceptions of self-concept intrinsically include efficacy expectations within the domain of interest (Bandura, 1986). One cannot accurately assess their standing within a domain without considering how they expect to perform on tasks within that domain.

Finally, a third difference between self-efficacy and self-concept is the social dimension of the two constructs. Ability judgments for self-concept are thought to be influenced by how one gauges the capabilities of his or her peers (Bong & Clark, 1999). However, Bong & Clark (1999) argue that this could be an artifact of the measures used for self-concept, some of which include items that set respondents up for social comparison. Nevertheless, several studies suggest that students differ in their self-concept based on how they are categorized as students, placed within their courses, and perceive their standing among their peers (Coleman & Fults, 1982; Marsh, 1986; Rogers, Smith, & Coleman, 1978). Bong and Clark (1999) also point out that self-concept is more evaluative than descriptive. A description of one’s capability could be stated as “I can perform this task.” An evaluation of one’s capability could be stated as “I can perform this task very well.” Inherent in the evaluation is a normative comparison. What is considered ‘very well’? One must judge one’s capability or competence against what is generally accepted to be ‘very well’. Because of this, performance scales used in studies of self-concept are typically standardized exams (Bong & Clark,

1999). Self-efficacy beliefs and the measures associated with them are less likely to trigger students to make self-comparisons with their peers. This is due to the nature of self-efficacy beliefs, which are most closely tied to direct experience with a specific task (Bandura, 1977). According to Bandura (1977,1986), mastery experience is the greatest determinant of one's efficacy appraisals. Thus, if measured appropriately, scores from a self-efficacy scale should yield an estimation of how respondents judge their personal capability to successfully complete a given task, as opposed to a self-concept score that is inherently tethered to social comparisons of ability. This difference is reflected in the predictive power of self-efficacy versus self-concept on academic achievement. In a recent meta-analysis of studies conducted with college students, self-efficacy was found to correlate much stronger with academic achievement than self-concept (Robbins et al., 2004).

Self-Efficacy and Academic Achievement

Self-efficacy is hypothesized to have far-reaching implications in academics by influencing students' effort, perseverance, and emotional reactions to specific tasks in school (Lent et al., 1984; Lopez & Lent, 1992; Pajares, 1996; Pajares & Kranzler, 1995). Moreover, self-efficacy has been positively associated with students' self-regulated strategies (Pintrich & De Groot, 1990; Zimmerman et al., 1992) and college major choice (Betz & Hackett, 1983; Lent et al., 1984; Lent & Hackett, 1987). Lent et al. (1984) investigated the relationship of college students' self-efficacy beliefs to persistence in technical and science majors. They found that students who reported higher self-efficacy scores for completing

their educational requirements were more likely to persist in their major. This supports Hackett and Betz's (1981) hypothesis that self-efficacy is linked to persistence in career goals.

One of the most widely studied effects of self-efficacy is the positive relationship it has with academic performance (Lightsey, 1999; Multon, Brown, & Lent, 1991; Robbins et al., 2004). Regardless of ability level, researchers have found that students who report high self-efficacy tend to outperform their peers who report low self-efficacy (Bouffard-Bouchard, Parent, & Larivee, 1991; Collins, 1982 as cited in Bandura, 1997). Schunk (1989) performed a series of studies with students who demonstrated severe deficits in mathematics and language skills. Instructional treatments were applied in an attempt to promote growth of self-efficacy. Students who showed the highest performance attainments were those who reported the highest self-efficacy, when compared with students who had acquired the same level of skills. These studies highlight that students with equal ability tend to perform at a level consistent with their efficacy expectations. This trend can also be seen among students at the college level as well.

Many studies have found that students' self-efficacy significantly predicts academic achievement at the college level (Lent, Brown, & Larkin, 1986; Pajares & Kranzler, 1995; Pajares & Miller, 1994, 1995; Siegel, Galassi, & Ware, 1985; Zusho et al., 2003). Multon, Brown, & Lent (1991) conducted a meta-analysis on studies that measured self-efficacy with samples of college students. They reported an average correlation of 0.38 for self-efficacy and academic performance, and that self-efficacy accounted for 14% and 12% of the observed

variance in academic performance and persistence, respectively. A subsequent meta analysis conducted by Robbins et al. (2004) surveyed 109 studies where various psychosocial and study skills variables were compared with academic performance. The two strongest psychosocial predictors of college GPA were self-efficacy and achievement motivation (ρ s = 0.496 and 0.303, respectively). Furthermore, due to the non-compulsory nature of college, many studies investigating self-efficacy among college-age students have examined the predictive power of self-efficacy on persistence (Hull-Blanks et al., 2005; Lent et al., 1984; Vuong, Brown-Welty, & Tracz, 2010). Wright, Jenkins-Guarnieri, & Murdock (2012) sampled 401 first-year undergraduates and found that course self-efficacy measured at the end of the semester was a significant predictor of persistence to enroll in the second semester. This effect was found after controlling for relevant variables such as gender, high school GPA, and ethnicity. Although the positive effects of self-efficacy on performance and persistence have been consistently reported for college as a general academic domain, the question remains: What are the effects of self-efficacy in specific disciplines, such as chemistry?

Self-Efficacy in College Chemistry

While motivation research was booming in other areas of academia, very little research on motivation occurred during the 1980s and 1990s in college-level chemistry. Studies addressing self-efficacy were no exception. The earliest studies to appear were unpublished, but demonstrated that self-efficacy was an

important construct tied to academic outcomes, and worth exploring further in the domain of chemistry (Kerns, 1981; Smist, 1993).

Much later, Dalgety, Jones, & Copolla (2003) published the first major instrument for measuring motivation-linked variables specific to college chemistry. The Chemistry Attitudes and Expectancies Questionnaire (CAEQ) contains 69 items and is comprised of three main scales, (self-efficacy, attitude toward chemistry, and chemistry learning experiences) with several underlying subscales. The self-efficacy scale, in particular, contains 17 items and was found to have no meaningful substructure due to high intercorrelations among proposed factors (Dalgety et al., 2003). The CAEQ, in full and in part, has been used in several studies since its development (Dalgety & Coll, 2006a, 2006b; Villafane, Garcia, & Lewis, 2014).

In the same year the CAEQ was published, Zusho, Pintrich & Coppola (2003) published a study that utilized a chemistry-adapted version of the Motivated Strategies and Learning Questionnaire (MSLQ), which contains a self-efficacy subscale. A sample of 458 undergraduate students enrolled in an introductory chemistry class were included in the study. The authors found in general, students' self-efficacy declined across the semester, supporting prior work which suggests that students' motivation tends to decrease with time (Schunk & Pintrich, 2002). High achievers were an exception, and actually reported higher self-efficacy at the end of the semester. Furthermore, self-efficacy was found to be the best predictor of course performance even after accounting for prior achievement. Although present in other physical sciences

and mathematics education research (Andrew, 1998; Bong, 2001; Lent et al., 1984), the relationship between achievement and self-efficacy is largely unexplored in college chemistry.

A very recent publication by Villafane, Garcia, & Lewis, (2014) focused on the trends of students' chemistry self-efficacy based on ethnicity and gender. Self-efficacy was measured five times throughout the semester of a preparatory chemistry course using five selected items from the CAEQ self-efficacy scale. All gender/ethnicity combinations exhibited an upward trend in self-efficacy, except Hispanic and Black males who showed a downward trend in self-efficacy across the semester. This study highlights that the expected outcome of increased confidence in chemistry following a preparatory course (Schmid, Youl, George, & Read, 2012) may not be true for all students.

The CAEQ remains the most widely cited measure of self-efficacy in college chemistry, though it has been used sparingly in actual studies. Another self-efficacy scale designed for college chemistry is the College Chemistry Self-efficacy Scale (CCSS) (Uzuntiryaki & Aydin, 2009). The CCSS contains 21 items in three subscales (self-efficacy for cognitive skills, psychomotor skills, and everyday applications). It was originally developed in Turkey, but was written and administered in English. The authors found that scores on the instrument were different for chemistry majors versus non-majors, with majors scoring higher. This is expected, as students majoring in chemistry should demonstrate more confidence toward the subject matter than non-majors. Since its development, the CCSS has been used in several university-level chemistry research studies

(Aydin, Uzuntiryaki, & Demirdög n, 2011; Uzuntiryaki, 2008; Uzuntiryaki & Capa-Aydin, 2013). None of these studies examined the relationship between self-efficacy and course performance, and all three were conducted in Turkey. To enhance the external validity of this instrument, studies in English-speaking countries are necessary (Uzuntiryaki & Aydin, 2009).

While self-efficacy is reputed as one of the strongest predictors of academic achievement, including persistence in the face of difficulty (Bandura, 1997; Pajares, 1996; Zimmerman, 2000), few researchers have realized its potential in college chemistry. Instead, chemical education researchers have focused on attitudes (Bauer, 2008; Xu & Lewis, 2011; Xu, Villafane, & Lewis, 2013), and self-concept (Bauer, 2005; Lewis, Shaw, Heitz, & Webster, 2009) as indicators of motivation. Although attitudes and self-concept are important variables to consider with chemistry students, self-efficacy should not be neglected when considering the motivation levels of students.

Interest

Interest, often referred to as personal interest or individual interest, is defined as “a relatively enduring predisposition to reengage particular contents over time” (Hidi & Renninger, 2006). Interest exists between a person and an activity (Deci, 1992). Hidi and Baird (1986) argued that interest is more than “arousal,” but must be considered as a process. As a process, interest is said to endure and persist through time. In the same vein as efficacy beliefs, interest is content-specific and represents a personal significance between the individual and the object of his or her interest (Renninger & Hidi, 2002; Schiefele, 1991).

Although commonplace in everyday language, the term “interest”, when considered as a psychological variable in educational studies, requires the same careful inspection and definition given to other psychological constructs. It is terms such as these which are easiest to misinterpret and misuse on account of implicit understanding.

Although prior work existed on interest (see Hidi, 1990 and Schiefele, 1991) John Dewey is the best-known author and philosopher who wrote on interest and education prior to modern interest theories. His book, entitled *Interest and Effort in Education*, contrasted learning based on interest and that based on effort (Dewey, 1913). He argued that learning based on effort lacked meaning for the individual, and resulted in trained knowledge (Schiefele, 1991). Dewey also introduced a theory of interest that was based on three tenets: interest is an active, “propulsive” state, it is based on real objects, and it has high personal meaning (Schiefele, 1991). Dewey’s work and theory of interest was incredibly influential and established the groundwork for modern interest theories and their connection to motivation.

During the era dominated by behaviorism, interest research was overshadowed by the more popular drive and mechanistic theories of motivation. However, the dawn of cognitive psychology in the 1960s provided an avenue for interest research to get off the ground again (Schraw, Flowerday, & Lehman, 2001). Motivation research that included personal interest re-emerged in Germany with the work of Hans Schiefele (Schiefele, 1974). He argued against his contemporaries who focused on performance-based motivation in education,

by highlighting the importance of the content to be learned and the value placed on it by the student. Dweck (1986) supported the shift away from a performance-only focus by arguing that performance-based goals tend to undermine intrinsic interest.

Once interest research began to take off in the 1980s and 1990s, two components and two conceptualizations of interest emerged. The two components of interest are feeling-related and value-related valences (Schiefele, 1991). Feeling-related valences are associated with positive (or negative) emotions and feelings of enjoyment. Value-related valences are linked to importance and a deeper, personal significance. The separation of these two components of interest has been confirmed empirically (Linnenbrink-Garcia et al., 2010). The two conceptualizations of interest are individual interest and situational interest (Hidi, 1990). Individual interest is defined above and is a more easily understood term than situational interest, which refers to interest that is generated through interactions with a stimulus and/or concrete objects under certain conditions of the environment (Hidi, 1990; Krapp, Hidi, & Renninger, 1992). Situational interest is said to be “triggered” by stimuli in the environment, and is typically accompanied by short-term changes in affect and cognitive processing (Hidi & Renninger, 2006). In contrast to individual interest, which exists internally, situational interest tends to be more of an external construct that is dependent on factors outside the person (Alexander & Jetton, 1996). However, Hidi and Baird (1986) caution against attributing interest to either an internal or an external factor. That is, interest cannot exist independent of the person nor

can it exist only within the person. It is a connection between the person and the stimulus, and depends on the stimulus as much as it depends on individual factors within the person. When interest has been sparked or triggered by an event, such as an engaging text or chemistry demo, this can mark the beginning of developing of a more enduring, persistent individual interest. Although the development of interest is not the same for all individuals, Hidi and Renninger (2006) argued that there is no evidence suggesting well-developed interest can spawn without the individuals first being exposed to the area and experiencing triggered interest.

The process by which well-developed individual interest is hypothesized to develop from triggered situational interest is outlined in Hidi and Renninger's (2006) four-phase model of interest development. This model describes the development of interest as progressing from situational interest to individual interest, a theory that is supported by several authors (Krapp et al., 1992; Renninger, 2000; Schraw et al., 2001; Silvia, 2001). However, it should be noted that the terminology used by authors is not always exactly what has become commonly accepted by the research community as "situational interest" and "individual interest". Other terms such as intrinsic interest or personal interest are used frequently, but these and other labels can be conceptualized in terms of situational interest and individual interest (Hidi & Renninger, 2006).

The four phases outlined by Hidi and Renninger (2006) in chronological order are: triggered situational interest, maintained situational interest, emerging individual interest, and well-developed individual interest. Triggered situational

interest refers to interest that is usually externally supported and is associated with a transient shift in affect and cognitive processing. Put simply, a person's interest can be piqued, or triggered, by an event that occurs to them not entirely by their own making. For instance, a student is in a physics class and the instructor discusses the potential of using magnets to generate electricity with a direct current motor. This student did not produce the event that triggered his interest, but was merely a passive observer. However, his cognitive processing and affective state changed as a result of the stated application. The application given by the instructor is an example of a "catch" factor, something that arouses attention, but may or may not stimulate the person to look further into the content.

Maintained situational interest is characterized by repeated interactions with specific content over time, but is still mostly externally supported. If the student mentioned above is now more attentive during lectures, asks questions, and even researches physics content having to do with the class on his own, he is demonstrating maintained situational interest. The triggered interest influenced his decision to assign more personal value to the content being taught in the class, as evidenced by his increased attention and attendance. If the way the class is taught or the type of content presented continues to be stimulating to that student, such factors would be referred to as "hold" components (Harackiewicz, Barron, Tauer, Carter, & Elliot, 2000). "Hold" interest differs from "catch" interest in that it tends to predict a more enduring interest due to the increased significance placed on the content by the person (Harackiewicz et al., 2000;

Mitchell, 1993). If a person is exhibiting maintained situational interest, his or her interest is “held” to some extent indicating he or she personally values the content and connects with it.

When someone seeks out opportunities to re-engage with content or opts to participate when an opportunity presents itself, this person has developed individual interest. Individual interest is exemplified when a person is resourceful, puts forth effort that feels “effortless”, and regularly generates his or her own curiosity on the subject (Flowerday & Schraw, 2003; Hidi, 1990; Hidi & Renninger, 2006). Emerging individual interest precedes well-developed individual interest, and tends to be somewhat externally supported and may require encouragement to persist through challenges (Hidi & Renninger, 2006; Renninger, 2000). In the case of the student from above, his interest in physics would be considered emerging individual if he started to seek out opportunities to learn and engage more with the content outside of class. By self-directing and supporting his growing interest in physics, he is showing the effects of valuing the content and assigning personal meaning to it. From there, his interest may or may not continue to advance toward a well-developed individual interest. If it does, the students’ interest in physics would likely persist beyond the semester, he would continue to seek out answers to questions based on curiosity, and his interest would persist in the face of difficulty (Hidi & Renninger, 2006; Renninger & Hidi, 2002). While there exists no delineated mark that separates emerging from well-developed interest, the degree to which one values, and the extent a person will go to engage the object of their interest are indicators of the

development level of their interest. Further, as interest deepens, the desire for knowledge and value for the content grow concurrently (Hidi & Renninger, 2006). Interest that endures beyond the situational “catch” is indicative of a personal involvement and significance that can be predictive of future choices in that domain (Mitchell, 1993). The implications of interest research are most resolutely pursued by those in fields of education.

Interest and Education

While interest is a psychological consideration in every domain from sports and careers to hobbies and culinary preferences, education remains the most visible outlet for interest research. Dewey (1913) was particularly invested in education and the lack of focus on students’ interest in the classroom. He argued for education to be more tailored to the students’ interests because he saw interest as an incredibly powerful motivator. His theories concerning interest and education were built upon, and ultimately became “most relevant for modern conceptualizations of interest” (Schiefele, 1991). Hence, interest research began in the field of education and, to a large extent, has remained there.

Interest has overwhelmingly been shown to positively affect learning in school (Asher, Hymel, & Wigfield, 1978; Estes & Vaughan, 1973; Frenzel, Goetz, Pekrun, & Watt, 2010; Harackiewicz et al., 2008; Pugh, Linnenbrink–Garcia, Koskey, Stewart, & Manzey, 2010; Schiefele, 1991). The focus of interest research in education has centered on how interest influences learning, how interest can be enhanced in education, and how it is linked to student motivation (Schiefele, 1991).

Several studies have investigated the role of interest in the quality of learning from text (Asher et al., 1978; Schiefele, 1999). Most of the findings suggest that students who demonstrate higher interest (or intrinsic motivation) toward the subject area have a deeper comprehension of the text than those who show lower interest (Benware & Deci, 1984; Bernstein, 1955; Fransson, 1977; Schiefele, 1992). Schiefele (1992) conducted a series of studies on the relationship between the level of text comprehension and topic interest. In one study, a sample of college students were asked to rate their expected interest toward a topic that a text passage covered. The students were then asked to read the passage and subsequently probed for their comprehension of that text. Students that reported higher levels of interest toward the topic demonstrated a deeper level of comprehension, such as comparing and applying aspects of the text to novel situations, than those who reported lower interest. However, the two groups did not differ in levels of rote memorization or concrete-type questions. In a second study, Shiefele (1992) confirmed the results from the first study and found that students with higher topic interest were better at discriminating between true and false statements about the text. Although interest is not particularly predictive of whether a student will remember discrete facts, it is linked with the overall quality of the learning as evidenced by deeper and more abstract comprehension.

As with any variable related to motivation in education, one of the most important considerations for educators and researchers has been the relationship of interest with academic achievement. Shiefele, Krapp, and Winteler (1992)

published a meta-analysis that examined the effect of interest on academic achievement. Sixteen studies were included in the analysis covering 20 years of research from 1965 to 1985. Only studies that measured subject-specific interest (such as physics) were included in the analysis. The authors found an average correlation of 0.31 for interest with achievement across the 16 studies. When considering specific subject areas, biology and literature were the only two subjects (out of seven) that had correlations below 0.30. In addition, the authors showed that the interest effect was no different for grade level groups 5-10 and 10-12. One major concern that was pointed out by the authors was the lack of causal ordering in the studies. Virtually all of the studies were strictly correlational, limiting the interpretation of the results to just the magnitude of the relationship between interest and achievement. Correlations cannot indicate the direction of the relationship or the order of the relationship. However, more recent studies have investigated the causal relationship between interest and academic achievement through the use of path modeling.

Harackiewicz et al. (2000) investigated the short and long-term effects of college students' goal orientations, interest, and performance in an introductory psychology class. Interest was measured in two scales: interest (hold), and enjoyment (catch). The authors utilized path modeling to predict the causal ordering of the tested variables. Their results showed that in the short-term (one semester), students who reported higher interest (hold) received higher grades in the course. Although this effect was significant, the authors point out that students had already received feedback from two exams prior to the

measurement of their interest. Hence, the notion that level of interest accounted for the performance level could be muddled by the timing of the measurements. Despite this, the finding is noteworthy and has been supported by other work. Hulleman et al. (2008) performed a similar study and found that the utility value component of interest (i.e. importance for future), but not the intrinsic component (i.e. enjoyment of class), predicted academic performance in college psychology. In this study, the interest was measured prior to any exams, eliminating the problem of timing for causal ordering. It should be noted that the interest (hold) scale from the study by Harackiewicz et al. (2000) also included items that assessed utility value. Hence, the extent to which one values the content they are studying seems to be predictive of academic performance, an idea supported by other scholars (Bandura, 1986; Wigfield & Cambria, 2010).

While much of the research on interest centers on primary and secondary school levels, an important implication of interest research has emerged at the college level: the types of courses taken by students. This is of particular importance at the college level because course choices play a major role in choice of major and career (Harackiewicz et al., 2000). Several studies lend support to the view that subject-specific interest is predictive of how many classes are taken (Harackiewicz, Barron, Tauer, & Elliot, 2002; Harackiewicz et al., 2000; Harackiewicz et al., 2008; Lent et al., 2001) and even major choice in that subject area (Harackiewicz et al., 2002; Harackiewicz et al., 2008; Lent et al., 2001; Lent, Brown, & Hackett, 1994; Malgwi, Howe, & Burnaby, 2005).

Harackiewicz et al. (2002) performed a study over the course of seven years that examined students' trajectory through college, including the number of psychology courses taken, major choice, and performance. In an introductory psychology course, their interest and goal orientations were measured. These, along with high school performance, and the aforementioned college outcomes were included in a path model to assess the predictive nature of the psychological constructs. The authors found that interest was a mediator between mastery goals and both, psychology courses taken and choice to major in psychology. Put simply, students who endorsed mastery goals were more likely to take more psychology courses and major in psychology due to an increased interest in psychology. The effect of interest on psychology courses taken was stronger for students with higher grades ($\beta = 0.43$) versus those with lower grades ($\beta = 0.21$) in the introductory psychology course. The same was true for the choice to major in psychology ($\beta = 0.33$, high grades; $\beta = 0.13$, low grades). Collectively, the research supports the expectation that a higher interest in a domain would confer the choice to major in that domain.

Interest as a Variable in Chemistry

As with most other psychological constructs, research that is concerned with the levels and effects of students' interest in chemistry is scarce. Considering the importance of interest on career and major choice and the urgency outlined by several reports to retain STEM majors (National Research Council, 2003, 2012; President's Council of Advisors on Science and Technology, 2012), it is surprising that the study of interest in chemistry has been

largely neglected. Nevertheless, a few studies have emerged that include interest as a key variable in the investigation. Due to the low number of studies on interest in chemistry, this review has been expanded to include high school chemistry in addition to college chemistry.

Nieswandt (2007) measured three chemistry-specific affective variables (self-concept, situational interest, and attitude), and examined their relationships with each other and with conceptual understanding in chemistry. The sample included students ($n = 73$) who were in ninth grade chemistry in Germany. The items used to measure situational interest included components of personal significance, enjoyment, and engagement. The personal significance items most closely aligned with the value-related valence outlined by Schiefele (1991). Also, the enjoyment and engagement items aligned closely with the feeling-related valence of interest. However, these components of interest were not separated in the statistical models, but were sub-factors of the latent trait situational interest. The affective variables were measured twice during grade 9. Conceptual understanding was measured twice during the year, at the end of grade 9 and start of grade 10, and aggregated together as one score. The author found four models that were tenable, but only two of them had a significant path coefficient to conceptual understanding. The first model suggested that situational interest at time 1 predicted situational interest at time 2 ($\beta = 0.51$), which then predicted conceptual understanding ($\beta = 0.37$). No other significant effects on conceptual understanding were found. The second model suggested that situational interest at time 1 predicted self-concept at time 2, which then predicted conceptual

understanding. Thus, the effect of situational interest at time 1 was mediated both by self-concept time 2 (model 2) and situational interest time 2 (model 1) in its effect on conceptual understanding. These results support the theory that interest is predictive of subsequent interest (Harackiewicz et al., 2008) and highlight that interest can influence self-concept in chemistry.

Dalgety and Coll (2006) investigated the factors that influence chemistry students' future course choices. The sample of students came from two semesters of introductory level chemistry courses. The psychological variables measured included attitudes, self-efficacy, career interest in chemistry, leisure interest in chemistry, and experiences in chemistry. These variables comprise the 69-item Chemistry Attitudes and Experiences Questionnaire (CAEQ), which was given to students three times over the course of two semesters - start of first semester, end of first semester, and end of second semester. The authors were interested in whether students differed on any of the scales based on whether or not they planned to enroll in a second year chemistry course. The largest difference on both career interest and leisure interest scales between these two student groups occurred at the end of semester two. Those who reported higher interest in chemistry were also those who intended on continuing with further education in chemistry. Again, this represents the connection between domain-specific interest and courses taken within that domain (Harackiewicz et al., 2002; Lent et al., 1994). Although the authors only measured students' intent to enroll, it is still valuable support added to the theory that interest predicts continuation in a field.

In addition to these two studies, there are several more that include interest as a minor component to the overall study in chemistry classes (Barbera et al., 2008; Price & Brooks, 2012; Salta & Koulougliotis, 2015). For example, Barbera et al. (2008), and Salta and Koulougliotis (2015) modified instruments for use in chemistry that had items related to personal interest. Barbera et al. (2008) found that students' interest in chemistry drops across a semester of instruction. Price and Brooks (2012) were interested in teachers' perceptions of their students' experiences following demonstrations in high school chemistry. Some of the instructors remarked that demonstrations increased their students' interest in chemistry, though there was no quantitative data to support this.

Interest can be a powerful motivator by directing adaptive learning strategies that will facilitate deep learning (Schiefele, 1991). Interest is also predictive of positive learning outcomes and the types of choices that students will make about their future (Krapp et al., 1992; Schiefele, Krapp, & Winteler, 1992). According to Hidi (1990), "interest is central in determining how we select and persist in processing certain types of information in preference to others" (p. 549). The state of research on interest in chemistry is lacking, and the field of chemical education could benefit greatly from investigating this psychological construct among its students. Importantly, by understanding more about what students enjoy and value in chemistry courses, instructors can better tailor the content of the course as a way to tap into the positive downstream effects of enhanced interest.

Effort Beliefs

Effort beliefs can be described as beliefs about the extent to which the expenditure of effort will lead to a desired outcome. The construct, effort beliefs, stems directly from implicit theories of intelligence, first outlined by Dweck (1986), but also has ties to attribution theory (Weiner, 1985). While implicit theories of intelligence are central to the understanding of effort beliefs, attribution theory is an important parallel theory to consider.

Brief Overview of Attribution Theory

Effort can be considered as part of the attribution theory of achievement, and from this theoretical standpoint, is intimately tied to conceptions of ability (Weiner, 1985). Attribution theory deals with the things to which people ascribe their successes and failures. First outlined in 1971, it was postulated that students attribute their achievement mainly to ability, effort, task difficulty, and luck (Weiner et al., 1971). Of these, Weiner (1985) argues that ability and effort are the most salient causal ascriptions to achievement. In short, students believe that those who have high ability and display high effort will be more successful than students who have low ability and display low effort. The notion that effort and persistence has a positive effect on the academic outcome of a student has been supported by empirical studies (Elliot, 1999; Stipek & Gralinski, 1996). For example, Elliot (1999) found that self-reported persistence and effort were positive predictors of academic performance. Effort was found to be a mediator between adaptive mastery goals and academic performance. Another study revealed that students who expressed positive beliefs toward the value of effort

do not necessarily show increased performance, but do tend to focus more on mastery and the development of their abilities (Stipek & Gralinski, 1996). Effort is certainly a key component in the academic success of students. It must be considered when making judgments about academic performance, due to how it mediates the link between motivational constructs and academic outcomes (Elliot, 1999; Goodman et al., 2011).

Implicit Theories of Intelligence

While attribution theory focuses on causal ascriptions for success and failure, the work underlying implicit theories of intelligence is concerned mainly with conceptions of ability (Dweck, 2002). This includes how conceptions of ability are developed, how they are shaped by individual circumstances and experiences, and how motivational processes stem from these conceptions.

By definition, implicit theories, in contrast to scientific theories, are not explicitly articulated in the mind of the beholder (Burnette et al., 2013). That is, people are not necessarily aware of the implicit theories they hold. Two major implicit theories of intelligence have been theorized: incremental and entity (Dweck, 2002). Someone who holds an incremental theory believes that intelligence can grow with effort; whereas someone with an entity view believes that intelligence is fixed and cannot change (Dweck & Leggett, 1988). Those who endorse an incremental theory of intelligence tend to be more optimistic and persistent in the face of difficulty, whereas those who endorse an entity theory are more likely to lose confidence and even give up when encountered with setbacks (Dweck, 2002). In addition, students with incremental theories are more

likely to adopt mastery (or learning) goals (Dweck, 2000; Dweck & Leggett, 1988), which signify the students' propensity to increase their competence (Dweck, 1986). Mastery goals have been shown to mediate the relationship between incremental theories and academic achievement (Blackwell et al., 2007). However, in east-Asian cultures, the relationship between incremental theories and academic achievement has been shown to be mediated by performance-approach goals (Chen & Wong, 2014). Performance goals, in contrast to mastery goals, are mainly characterized by a focus on looking smart and outperforming peers, and less on developing competence (Dweck, 2000; Dweck & Leggett, 1988). Regardless of the mediator, students who adopt incremental theories are more likely to outperform their fellow classmates who adopt an entity theory of intelligence. Dweck (2002) argues that these effects are evident because students with incremental views display increased effort and see challenges as a part of the mastery process. On the other hand, entity theorists view setbacks as a threat to their fixed intelligence, and may be defensive and avoid future challenges.

The work dealing with implicit theories has largely focused on lower grade levels (elementary, middle), perhaps due in large part to the fact that most children develop conceptions of ability and competence at an early age (Dweck, 2002; Wigfield et al., 1997). However, more recent research has started moving up age groups toward high school and particularly, the college level.

In a study on college students at the University of California at Berkeley, Robins and Pals (2002) were interested in the predictive power of implicit

theories of intelligence. They examined achievement goals, mastery versus helpless orientations, self-esteem, and grades over the course of four years. Although the entity theorists came to college with higher grades, this did not translate into higher achievement relative to the incremental theorists. Further, entity theorists were more likely to attribute setbacks to their lack of ability, and on average declined more in their self-esteem across their four years in college relative to the incremental theorists. Thus, the entity theorists tended to be more vulnerable in academic situations than the incremental theorists. Despite this, the two groups did not differ on academic achievement, indicating that implicit theories did not predict achievement directly. Others have also failed to establish a direct link between implicit theories and academic achievement in college courses (Shively & Ryan, 2013).

Although Dweck (2000) argues that implicit theories of intelligence are foundational and absolutely vital for shaping one's academic achievement, the relationship between implicit theories and academic performance has only been demonstrated indirectly for college students. For example, in a study performed with Chinese university students, researchers found that incremental theories predicted performance-approach goals, which then predicted academic GPA (Chen & Wong, 2014). Mastery goals were also a mediator between incremental theories and academic performance, but the relationship was much weaker. This stands in contrast to similar Western studies, which have found that incremental theorists are more likely to adopt mastery goals (Blackwell et al., 2007; Dweck & Leggett, 1988). This discrepancy could be due to cultural differences and the

structure of Chinese education that is so focused on high stakes standardized testing (Chen & Wong, 2014).

Parallel to conceptions of intelligence are conceptions about effort. As Dweck (2000) points out, effort has different meanings for entity theorists and incremental theorists. Entity theorists view effort as an indicator of intelligence, where more effort signifies less intelligence. In contrast, incremental theorists view effort as the path to their growing intelligence. Thus, when entity theorists encounter an obstacle that cannot be overcome without high effort, they may call into question their intelligence levels. The overlap between effort beliefs and implicit theories is strong, but has only been demonstrated in a few studies.

Prior Research and Implications of Effort Beliefs

There exist very few published educational studies that include a measurement of effort beliefs at any grade level. As a result of the paucity of research relating to this construct, the following review of literature will include studies from all grade levels and domains. The earliest studies on effort beliefs added a dimension to implicit theories of intelligence, by using effort beliefs to further differentiate entity theorists from incremental theorists (Dweck, 2000). One study showed that entity theorists, more so than incremental theorists, endorsed statements such as, “You only know you’re good at something when it comes easily to you” (Leggett & Dweck, 1986, as cited in Dweck, 2000). Similar findings were also reported for college-aged students, whereby entity theorists agreed with statements like, “If you’re good at something, you shouldn’t have to work very hard to do well in that area”, significantly more than incremental

theorists (Mueller and Dweck, 1997, as reported in Dweck, 2000). These two studies, although informative, only establish the relationship between implicit theories and effort beliefs.

In an attempt to extend the framework of implicit theories, effort beliefs, and goal orientation, a study by Stipek and Gralinski (1996) was conducted. Not only were several variables included, but also the domains of math and social studies were targeted specifically. This was a novel approach for implicit theories, as most of the research prior to that had targeted intelligence beliefs for school in general. Several motivational variables were measured among elementary school students, including effort beliefs, implicit theories, goal orientation, and performance. The authors found significant correlations between positive effort beliefs (e.g., “Everyone could be smart in math if they worked hard.”) and mastery orientation (e.g., “I do my math work because I like learning new things.”). As expected, positive effort beliefs did not correlate with entity-related beliefs (e.g., “Some kids can’t do well in any kind of school work.”), a finding has been confirmed in another study as well (Abdullah, 2008). However, positive effort beliefs did not correlate with academic performance either, indicating that even if students believe they can be good at math with hard work, they still may not achieve high marks. However, students endorsing entity-related beliefs did perform below their peers who did not endorse them, revealing that a student who believes he or she will never be good at math will likely underperform in math.

A few more recent studies have looked at the meditational role of effort beliefs in relating implicit theories with positive learning strategies and performance. Blackwell, Trzesniewski, & Dweck (2007) investigated how implicit theories related to academic performance among seventh and eighth grade students. The authors were not only interested in finding a relationship between implicit theories and performance, but were also focused on the reasons why such a relationship exists. Toward that end, they tested other variables in their model, including positive effort beliefs, positive strategies, helpless attributions, and learning goals. To measure students' helpless attributions and positive strategies, they were given a fictional scenario in which the student likes the material pretty well and studies a "medium amount" for the first quiz only to get it back with a grade of 54. The prompt then asked students how they would respond to this situation with a series of items related to low helpless attributions (e.g., "I didn't study hard enough.") and positive strategies (e.g., "I would work harder in this class from now on."). The findings suggested that positive effort beliefs were most highly correlated with positive strategies and low helpless attributions. Incremental theory was also strongly correlated with effort beliefs. None of the variables were significantly correlated with sixth grade performance, showing that the students' positive motivational dispositions were not related to prior achievement. Plotted performance trajectories for seventh and eighth grade math achievement displayed that students who endorsed an incremental theory showed gains in achievement, while entity theorists showed losses. A causal path model that included all of the aforementioned variables revealed that implicit

theories predicted positive effort beliefs, which predicted low helpless attributions and positive strategies. Both of the latter variables predicted increasing math grades across junior high school. Thus, effort beliefs is crucial as a mediator between incremental theory and positive strategies, as well as between incremental theory and low helpless attributions.

A study carried out by Jones and colleagues (2012) attempted to replicate Blackwell et al. (2007) model among ninth grade math students. The authors included a few key differences from Blackwell et al. study. Their study was directed at ninth grade math students using domain-specific measures; they measured students' interest, and they assessed the impact of achievement on the model both as a predictor and an outcome. The model and path coefficients were very similar in magnitude and direction to those found in the Blackwell et al. study, with a few exceptions in magnitude only. Positive effort beliefs were found to influence both positive strategies and low helpless attributions, but were also significantly influenced by current math grade. This suggests that students who are already performing well in math are reinforced to believe that their effort is fruitful. In a separate model, where current math grade was an outcome, interest was found to be positively influenced by incremental theory, and predicted learning goals. The effect of interest on learning goals is expected, as someone interested in math is more likely to want to learn about it, but the effect of incremental theory on interest was surprising to the authors. The authors offered the possibility that both interest and incremental theory might be related to another unmeasured construct, such as general intelligence. In any case, the

findings of this study further expand upon the importance of considering effort beliefs as a mediator between implicit theories and adaptive learning patterns that lead to higher achievement (Tempelaar et al., 2015).

Although effort beliefs, as a construct, have not been studied much, they still hold important implications for educational research and practice. Dweck (2000) states, “Effort is one of the things that gives meaning to life. Effort means that you care about something, that something is important to you and you are willing to work for it” (p. 41). The landscape is wide open for researching effort beliefs, as suggested by the quote from Tempelaar et al. (2014),

Given that the large majority of empirical studies based on self-theories framework have *not* explicitly conceptualized effort belief constructs, we assert that the full potential of self-theoretical frameworks is yet to be achieved (p. 116).

Effort is a complicated term and can have positive and detrimental effects on school achievement. Students who try hard and achieve high marks are reinforced by their effort to continue exerting effort in the future. However, students who fail continually following very diligent attempts at success may be turned off to the idea of exerting effort in the future. As a result, effort has been described as a “double edged sword” (Covington & Omelich, 1979). Every student comes to a class with their own beliefs about, and experiences with, effortful actions. By studying how students form their conceptions about effort, how they can change, and how they relate to other motivational constructs, instructors will be in a better position to direct students toward adaptive learning patterns.

Connections Between Effort Beliefs, Interest, and Self-Efficacy

Self-efficacy, interest, and effort beliefs represent three distinct psychological constructs that can have an effect on motivation. To date, no studies have been found that link all three constructs together, either empirically, or theoretically. However, there have been several empirical studies that have included two of the three constructs as measured variables. Other authors have offered theoretical hypotheses for how these constructs might be related. This section of the review will focus on studies that have investigated connections between self-efficacy, interest, and effort beliefs.

Lent, Brown, and Hackett (1994) applied social cognitive theory to career development by formulating a model that included self-efficacy, academic interest, choice, and performance. Central to their model was the notion that self-efficacy is a major mediator of choice and development, and guides one's decision-making. In the authors' words, self-efficacy represents "personal convictions about one's generative capabilities to negotiate specific task or situational challenges" (p. 86). Self-efficacy, they argue, is tied to the development of interest. People's interests are reinforced by their performance accomplishments, which in turn, is a source of self-efficacy. Thus, the way in which one perceives oneself to be capable at a task or activity will influence their interest in that activity. This is not a static process, but a dynamic feedback loop in the early stages of interest development, whereby interest can spur someone to engage in an activity. Their self-efficacy is influential toward future engagements and goal setting, and also serves to reshape and redefine their

interest. People tend to be attracted to domains and activities in which they feel most efficacious (Bandura, 1986). If a person lacks confidence, his or her interest will likely diminish as a result. As the authors state, “it may be difficult for robust interests to blossom where self-efficacy is weak or where neutral or negative outcomes are foreseen” (p. 89). These notions are not purely theoretical, but are backed up by several empirical studies that have centered on the role of self-efficacy in students’ persistence toward science, engineering, and math degrees (Betz & Hackett, 1983; Campbell & Hackett, 1986; Hackett, 1985; Lent et al., 1984). Overall, these studies support the hypothesis that self-efficacy is linked to interest, in terms of college major choice. In their publication, Lent, Brown, and Hackett (1994) averaged correlations from all relevant studies available to them at the time of several motivational variables, including self-efficacy, interest, and performance. Of all the correlations with self-efficacy, interest was the strongest ($r = 0.53$), and performance was moderately correlated ($r = 0.38$). More recent studies have corroborated these findings (Larson, Stephen, Bonitz, & Wu, 2014; Lee, Lee, & Bong, 2014; Lent et al., 2001; Lent et al., 2008; Smith & Fouad, 1999). Based on the evidence presented here, it is suggested that interest and self-efficacy are interconnected and strongly correlated. The causal relationship between the two is likely to be difficult to deconstruct, though it has been suggested that self-efficacy is a causal precursor to interest (Lent et al. 2001). Still others have models of interest and self-efficacy that are temporally equivalent, with both being caused by variables not included in the model (Lee et al., 2014). In either case, these studies highlight the salience of considering

interest and self-efficacy in achievement motivation models, as well as career and college major choice models.

The empirical research on effort beliefs remains sparse, and as a result, there is little evidence that addresses connections with other motivational constructs. Nevertheless, a few studies do exist that have measured self-efficacy or interest in combination with effort beliefs (Abdullah, 2008; Jones et al., 2012). Abdullah (2008) found a strong, positive correlation ($r = 0.51$) between positive effort beliefs and self-efficacy with a sample of students. This is consistent with Bandura's (1997) prevailing theoretical model which suggests people with high self-efficacy tend to display more effort and persist longer with tasks than those with low self-efficacy. Although a display of effort is not equivalent to believing that effort will lead to positive outcomes, certainly positive effort beliefs toward a task must precede the exertion of effort. Someone who believes they can successfully complete a task will likely also believe that their effortful actions toward that task will lead to positive outcomes. Dweck (2000) presents a slightly different angle on the relationship between incremental theory of intelligence (related to positive effort beliefs) and confidence (self-efficacy). She argues that confidence, while a valuable asset and predictor of achievement when things are going well, is not sufficient to carry students through difficult transition periods during their academic years (e.g., transition to junior high, or college). Rather, students endorsing an incremental theory of intelligence are more likely to persist in difficulty, whether they have high or low confidence in their current ability or intelligence. On the other hand, those with an entity theory and low confidence

are more likely to lose ground when faced with obstacles by blaming their fixed intelligence rather than effort (Hong, Chiu, & Dweck, 1995). Although not a direct connection between effort beliefs and self-efficacy, this argument presents an alternative perspective of the grounding effect that an incremental theory of intelligence is thought to have on students with both high and low confidence.

The relationship between interest and effort beliefs was investigated in one study on ninth grade math students (Jones et al., 2012). The authors were interested in the impact of effort beliefs, mastery goals, positive strategies, interest, implicit theories, and low helpless attributions on each other and on current math grade. In a different model, current grade was used as a predictor of other variables. Although the causal models tested in the study showed no significant relationship between interest and effort beliefs, data were presented demonstrating a moderate, positive correlation between effort beliefs and interest. Effort beliefs and interest were, however, both predicted by incremental theories of intelligence in one of the causal models tested. This suggests that math students who believe their intelligence is malleable and can grow with effort tend to exhibit more interest toward math and also believe that their effort will lead to positive outcomes.

Although some research exists on the connectivity between self-efficacy, interest, and effort beliefs, it is largely fragmented. Even in these cases, most of the research seems to center on primary and secondary grade levels. All three variables have important implications for learning and motivation, but research that includes all three in a college setting remains to be completed. By

understanding more about the impact of self-efficacy, interest, and effort beliefs on each other and on course performance, researchers and educational practitioners can better address changes and refinements in the classroom to promote positive motivational patterns. It is these classroom innovations, whether major overhauls or brief interventions, that are the critical next step to connecting the theoretical formulations and empirical measurements with practical improvements to teaching.

Social-Psychological Interventions in Education

It is no easy task for teachers to succeed in having their students learn what is expected of them in a given course. This is complicated by the fact that every individual comes to the course with his or her own personal history. Factors such as cultural background, family life, learning disabilities, and prior experience, just to name a few, are among those that can potentially inhibit learning. There is no silver bullet or single approach to teaching that can accommodate every individual in the class. In spite of this fact, large-scale efforts have been made to overhaul the way teaching is practiced from elementary grades through college. Other researchers have taken a different approach to trying to influence student learning and motivation – social-psychological interventions. Typically, these interventions do not teach material related to the class. Instead, they target students' feelings, attitudes, and motivations toward the class, or academics in general (Yeager & Walton, 2011). Although countless interventions have been attempted over the years, only those which are salient and most relevant to this dissertation study will be reviewed. Due to the

contextual constraints (e.g., college-level, limited lecture time) imposed by this study, only interventions that are age appropriate, relatively brief, and adaptable to this context will be considered as relevant. Several of these types of interventions have been employed and published, some with remarkable results. However, no evidence was found that these or any other social-psychological interventions have been used in the context of chemistry education. Despite this, any of the following interventions could be adapted for use in a chemistry setting because they are not content-driven, but rather target students' experiences in school from the student's perspective (Yeager & Walton, 2011).

In two recent studies, a series of short writing assignments proved to be very impactful at reducing the gender-achievement gap in college physics (Miyake et al., 2010) reducing the racial-achievement gap among seventh grade students (Cohen et al., 2006). Both studies employed a similar procedure with a double blind experimental design, whereby students were randomly assigned to either the treatment or control group. The treatment groups were asked to indicate the two or three values from a series of choices (e.g., being good at art, religious values) that were *most* important to them. The control groups were prompted to select two or three values from the same list that were *least* important to them. On the back page, the treatment groups were then asked to reflect on the most important value they chose, and write a short paragraph about why that value was important to *them*. The control groups were asked to reflect on the least important value they chose, and then describe a situation when that value would be important to *someone else*. Cohen et al. (2006)

reported roughly a 40% decrease in the racial-achievement gap for the treatment group. Essentially, the difference in grade points between European Americans and African American students was 40% less in the treatment group than in the control group. No treatment effect was found for the European Americans, and the effect for African Americans was still observed months after the treatment was implemented. Moreover, a follow-up study revealed that the treatment effect persisted for low achieving African Americans two years later (Cohen, Garcia, Purdie-Vaughns, Apfel, & Brzustoski, 2009).

Similarly, Miyake et al. (2010) detected a reduction in the gender-achievement gap for their sample of undergraduate physics students. Women in the treatment condition performed better overall on exams and a higher percentage of women obtained B's in the course compared with the control condition. In addition, the gap between men and women in the control condition was statistically significant, while in the treatment condition it was not. The values affirmation treatment was particularly effective if the women endorsed the gender stereotype by agreeing with the statement, "According to my personal beliefs, I expect men to generally to do better in physics than women." Those in the treatment condition that endorsed this statement performed significantly better on exams than those in the control condition.

The question then arises: How can such a brief, seemingly irrelevant intervention have a measurable effect on the performance of one group relative to another? As Cohen and Garcia (2008) explain, "people want and need to see themselves in a positive light." Through the process of reflecting on personal

values, the students engaged in self-affirmation (Steele, 1988). By self-affirming, people are better able to tolerate a threat to their identity in some other domain (Cohen & Garcia, 2008). The long-lasting effects of the treatment are possible because the self-affirmation exercise can affect a student's experience in school, and set in motion recursive processes that swell in effect over time (Yeager & Walton, 2011). Students who begin to have a sense of belonging in school, or in a particular course because they were asked to write about something important to *them* may start to display stronger patterns of adaptive learning. Over time, these patterns can build and the positive results from these patterns feeds back producing a lasting effect.

Another type of educational intervention that has gained some popularity recently addresses students' implicit theories of intelligence. Implicit theories are self-beliefs, and these beliefs can have tremendous effects on how people function, in particular, how they learn (Dweck, 2008). A few studies have shown that students' implicit theories can be manipulated through reading a persuasive text passage or "scientific article" claiming that either incremental theory or entity theory has been demonstrated as the correct viewpoint (Hong, Chiu, Dweck, Lin, & Wan, 1999; Kray & Haselhuhn, 2007). However, these treatment effects were not studied with relation to academic achievement or for long-term persistence as the students were debriefed about the exercise following the manipulation. What these studies *do* illustrate is that students' mindsets can be changed relative to their implicit theories of intelligence.

Blackwell, Trzesniewski, and Dweck (2007) performed a more lasting intervention directed at teaching students about the malleability of intelligence and the physiology of brain connections related to intelligence. The intervention was a series of eight 25-minute workshops for seventh grade students. The students involved in this study were low achieving, mostly minority students, who on average, scored in the 35th percentile nationally on the sixth grade standardized math exam. Those in the treatment condition learned about the malleability of intelligence and how learning and effort establish new and stronger connections in the brain, which causes intelligence to grow. Students in the control section also learned about the physiology of the brain, but they were taught study skills in place of a lesson on the malleability of intelligence. Both groups were assigned randomly to either the treatment or control from pre-existing advisory classes to which the students were assigned randomly by the school at the start of the year. Based on prior math grades and prior measurements of the motivational variables, students did not differ significantly between treatment and control at the start of the intervention. Prior to the intervention, both control and treatment groups were showing a downward trajectory in their math grades from spring of their 6th grade year through fall of their 7th grade year. However, following the intervention in the spring of their 7th grade year, students who were taught about the malleability of intelligence displayed a remarkable upturn in their math grades. At the same time, those in the control group maintained their declining grade trajectory. This effect was remarkable and could have easily gone undetected for two major reasons. First,

the students had already received feedback on their status in math class by the time of the intervention, which occurred more than a third of the way into the spring. Second, the treatment and control workshops were very similar in content except for the few sessions on the malleability of intelligence. This study is further evidence that short interventions, repeated in this case, can affect the way students psychologically construct their worlds, and in this way, have real consequences for achievement.

Hulleman and Harackiewicz (2009) designed an experimental study involving high school science students and investigating whether attaching personal relevance to topics could increase students' interest and performance. Two groups were randomly assigned to a treatment (relevance) or control (topic summary) condition. The study took place throughout the entirety of a ninth grade semester, with the interventions embedded during that time. Students were given a notebook with instructions for either the control or treatment condition, and asked to write essays about the topics they were learning. Students in the treatment condition were asked to write how the topic was personally relevant to them, whereas those in the control condition were asked to write a summary of the material for that topic. On average, the students wrote 4-5 essays throughout the semester. The authors hypothesized that students who had low expectations for success at the beginning of the course would benefit most through the intervention. Indeed, their hypothesis was supported, and students with low expectations for success in the treatment condition displayed markedly higher grades (nearly two-thirds of a grade point) than those with low

expectations for success in the control condition. A similar trend, but lower in magnitude, was observed for students' interest in science. The same authors repeated a very similar intervention with college psychology students, by instructing students to write about aspects of the course that had relevance in everyday life (Hulleman, Godes, Hendricks, & Harackiewicz, 2010). Students with poor performance histories in the class and low perceived competence benefited the most by the randomized controlled intervention by increasing their perceived task value of the course. These studies demonstrate the potential positive effects that can be obtained from a simple intervention by encouraging students to find meaning and value in what they are learning.

Social-psychological interventions have shown much promise for improving academic achievement and reducing stereotype threats in several educational settings. In general, these interventions are not tied to the specific content of the course, but they certainly could engage students with the subject material (Hulleman & Harackiewicz, 2009). Although the studies reviewed here all had at least some measurable treatment effect, many other randomized control trials in education are unsuccessful at detecting any effects (see Yeager & Walton, 2011). Success aside, there is another common thread linking the aforementioned studies. Every one contains an element of stealth with regard to the interventions employed. This is accomplished by disconnecting the intervention from the effect being tested. For instance, the values affirmations given by Miyake et al. (2010) and Cohen et al. (2006), on the surface, had nothing to do with stereotype threat. Similarly, the workshops targeting implicit

theories of intelligence had little to do with math skills (Blackwell et al., 2007) and the relevance writing assignments were not geared toward mastering content knowledge in science (Hulleman & Harackiewicz, 2009). Yeager and Walton (2011) argue that indirect interventions may be more effective than overt strategies. In their words, “They [indirect approaches]...allow students to take credit for their success rather than risking the possibility that students attribute positive outcomes to a heavy-handed intervention” (p. 284). The simplicity and “stealthy” nature of the interventions listed here make them attractive candidates for instructors with limited resources. Although the 8-week workshops reported in Blackwell et al. (2007) would be difficult to manage for most instructors, these could be shortened and incorporated into one or several lecture periods. Even very short interventions that powerfully convey psychological ideas that promote adaptive learning patterns can be effective, both in short and long-term outcomes (Yeager & Walton, 2011).

Psychometrics

Tests, in a psychological sense, are generally given to measure some psychological attribute of a person. Psychometrics is an area of study that is chiefly concerned with the measurement of psychological attributes. Examples of psychological attributes include: intelligence, aptitude, depression, motivation, anxiety, and many others. Where testing deals with the attributes of a person, psychometrics deals with the attributes of the test (Furr & Bacharach, 2013). If the attributes of people are important enough to measure with a test, then the attributes of the test should be thoughtfully considered and systematically

evaluated. As Furr and Bacharach (2013) put it, “If something is not measured or is not measured well, then it cannot be studied with any scientific validity.”

Perhaps the two most cited and discussed “attributes” of tests are validity and reliability. This is because quality tests should produce data that can be interpreted accurately (validity) and consistently (reliability). However, the scope of psychometrics, while centrally connected to validity and reliability, extends to other “attributes” such as scaling, dimensionality, and bias (Furr & Bacharach, 2013). This section of the review will focus on validity, reliability, and the dimensionality of tests.

Two types of variables can be measured on a psychological test – those which are observed, and those which are not. Examples of observed variables could be performance tasks, such as a spatial ability test, or behavior observations, such as solving a Rubik’s cube. Unobserved psychological variables are referred to as constructs, because they cannot be measured directly (Murphy & Davidshofer, 2005). Interest and self-efficacy are constructs that cannot be directly quantified, but must be inferred based on some other observed variable(s), such as a self-report test. Unobserved variables, such as these, are referred to as latent variables by statisticians (Wagner, Kantor, & Piasta, 2010). The observed variables on which the latent variable is based are called indicators, and because the latent variable cannot be directly observed, the estimation of its value is subject to measurement error (Brown, 2006; Wagner et al., 2010). Indicators in survey research, for example, are the items on the survey. These items may indicate intelligence, for instance, through tasks such

as mathematical computations, or a psychological state by asking respondents to indicate their beliefs and feelings about something. Motivation and its related constructs are latent variables which are almost entirely assessed by means of self-report surveys, and survey research hinges on the connections between indicators and latent variables.

Test Dimensionality and Factor Analysis

The dimensionality of a test is a critical component to interpreting scores from that test. Some tests are intended to be unidimensional and are interpreted as a total score for one psychological attribute. Other tests are multidimensional, measuring more than one psychological attribute. For multidimensional tests, the scores for different attributes are usually interpreted separately (Furr & Bacharach, 2013). For example, suppose a science aptitude test is given to a sample of university faculty members, who are said to be “scientists”. On the test are items relating to astronomy, medicine, and chemistry. If one of the faculty members in that group scores a 30 out of 70 on the test, is his or her knowledge said to be scientifically inadequate, even if he or she got 100% of the chemistry questions correct? In this case, if the researchers neglected the dimensionality of the test, then the results would lack meaning. Science, like intelligence or motivation, covers many domains and cannot be understood as unidimensional. Thus it follows, a science aptitude test that covers separate domains of science cannot be interpreted as a single score of science aptitude. This example illustrates the importance of considering the underlying dimensions in a test. Is the test measuring one construct or several? If several, which items on the test

are measuring which construct? Finally, are the items measuring *only* the intended construct? Questions such as these can be addressed by evaluating the dimensionality of a test. By far, the most common empirical method employed to investigate the dimensionality of tests is factor analysis.

Factor analysis is a group of methods that are designed to determine the number of factors (latent variables) which account for the variance and covariance among a set of indicators (items) (Brown, 2006). That is, factor analysis is based on the variation between individuals and the underlying associations that exist between indicators. Clusters of correlated indicators are referred to as factors (Martella, Nelson, Morgan, & Marchand-Martella, 2013) and the degree to which the value of an indicator can be described by the factor is called the factor loading (Brown, 2006). If anxiety were a factor on a particular survey, then the level of one's anxiety at the time of the test would be expected to direct their answer choices to items dealing with anxiety. Items with the highest factor loadings are the items which are most closely correlated with the latent variable. Squaring the factor loadings will tell the researcher what percentage of the variance in the indicator is explained by the latent factor (Brown, 2006). For example, suppose an indicator of depression has a factor loading of 0.78. This means that $(0.78)^2$, or 60.8% of the variance in that indicator is explained by depression. The remaining variance is called unique variance, and is a combination of specific factor and measurement error variance (Brown, 2006).

Of course, all of this is predicated on the assumption that the indicators were chosen based on the underlying theory that describes the latent variable of

interest. One cannot indiscriminately assign a latent variable to a grouping of items without some theoretical foundation to support the content of the items (Furr & Bacharach, 2013).

Factor analysis can be grouped into two main types: exploratory and confirmatory. Exploratory factor analysis (EFA) is data-driven and conducted with no prior hypothesis on how many factors exist, or which items are associated with a common factor (Welch, 2010). Typically when a test is first being developed, data from a pilot study will be used in an EFA. An EFA utilizes the intercorrelations among indicators to generate a factor structure, which then informs the researcher as to which indicators cluster together to form common factors (Brown, 2006). Confirmatory factor analysis (CFA), as the name implies, is a confirmatory method, whereby a factor structure is already hypothesized and the output generated tells the researcher how well the data fit the hypothesized model. CFA generally follows EFA, where a model has already been generated and can be tested again with an independent sample. Both EFA and CFA are an integral part in generating evidence for validity during survey development. Once a survey has been shown to generate valid data for a given population, CFA can be used again with scores from a different population to assess its functionality in that context. As will be discussed, validity is the major concern when interpreting test data, and valid interpretations depend on many factors, including the population from which the sample was drawn.

Validity

Validity is perhaps most easily described as, “the meaning of the test scores” (Messick, 1995). Validity has often been mistaken as a property of a test, but in fact, validity is not a property of a test or of its scores (Furr & Bacharach, 2013). Rather, validity is “the degree to which evidence and theory support the interpretations of test scores entailed by the proposed uses” (AERA, APA, NCME, 1999, p. 9). Consider, for instance, a ruler. In and of itself, a ruler cannot be considered a valid measurement tool of length. It must be used in accordance with how it will generate accurate measurements. One cannot put a ruler up to the moon and say the moon measures 0.5 inches across with any scientific credibility, no more than a “valid” survey on eating habits can be given to a monkey and produce believable scores. However, even if someone knows how to properly use a ruler, it does not mean his or her conclusions are valid. One must understand what length is and how to interpret the measurement obtained. If a room is found to be 13 feet in length, and the individual remarks, “That’s a heavy room”, then he does not have an understanding of what length means. The same applies for the measurement of psychological attributes. Just because a survey has been successfully used by others to draw valid conclusions does not make it a valid survey. It must be used in conjunction with how it was originally intended, and the researcher must know how to interpret the findings from the data produced.

In the past, four distinct types of validity evidence – content validity, construct validity, predictive validity, and concurrent validity - were proposed as

the “four faces of validity”, and could be emphasized to varying degrees depending on the purpose of measurement (APA, AERA, NCME, 1954; Loevinger, 1957; Murphy & Davidshofer, 2005). Tests in education, for instance, may have focused more heavily on content validity and predictive validity, as evidenced by expert opinions of test content in a domain and correlations to success in school, respectively. However, Messick (1995) argues this is an inadequate approach to evaluating a test. Instead, he contends for a comprehensive definition of construct validity as the umbrella term that encompasses all other branches of validity. By definition, construct validity deals with the extent to which the interpretations of scores from a test, given under specific conditions, represent a particular psychological attribute. The view that construct validity is central to overall test validity has become the contemporary perspective in recent decades, as evidenced by the updated *Standards for educational and psychological testing* (AERA, APA, NCME, 1999). As a result, all other evidences of validity are subsumed beneath construct validity, because the meaning and interpretation of test scores in specific situations are the central concerns of all aspects of validity (see Figure 1).

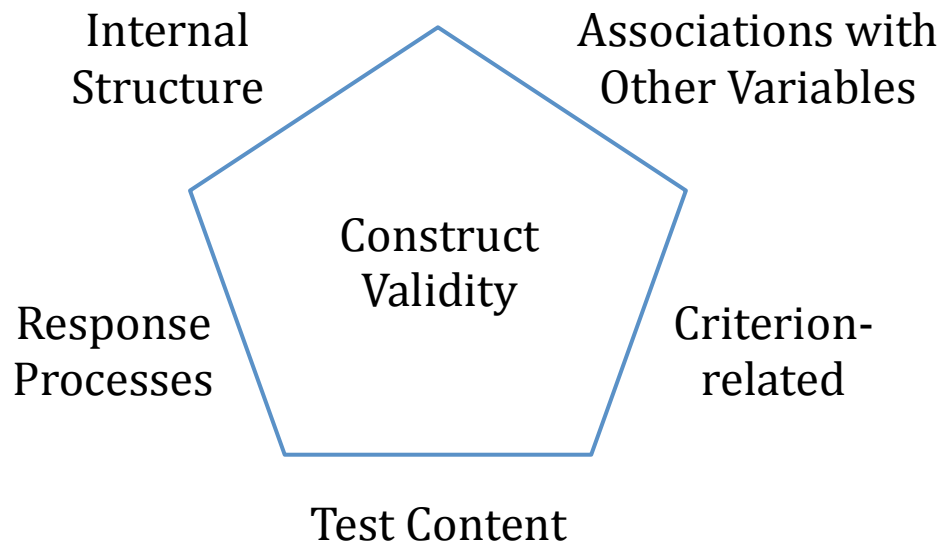


Figure 1. Representation of the contemporary viewpoint of validity evidences for testing purposes.

Evidence for test content validity. Every test is designed to measure some psychological attribute, or construct; and because a construct is an abstraction of the mind, it must be reified into select behaviors that are representative of the construct (Martella et al., 2013). Behavior implies specific concrete action and tasks, as many tests aim to assess, but in the same way, responding to self-report surveys about one's beliefs could be considered metacognitive behavior. There are behaviors which represent the construct (relevant behaviors) and there are those which do not represent it (irrelevant behaviors). The degree to which a test measures construct-relevant behaviors and leaves out construct-irrelevant behaviors is evidence of its content validity (Furr & Bacharach, 2013; Murphy & Davidshofer, 2005).

Evidence for response process validity. When people are given a test or survey, they engage in certain psychological processes to assign answers to

the questions. It may be assumed by the test giver that all participants utilize similar processes, and thus, any differences observed between individuals are true differences in the attribute being measured. However, the cognitive processes that people use to interpret and answer a given question are not always homogenous. For example, if a question were posed to chemistry students, "How confident are you to tutor a fellow student in this class?" The word "tutor" in this case could be interpreted by one student as simply helping a fellow student on a few homework problems. Or, perhaps a different interpretation might come from a student who has been tutored before in another subject by an expert. The second student might feel that he or she needs to have a substantial expertise in chemistry before feeling confident to tutor a fellow student. Questions like these should be subjected to a thorough investigation of the underlying cognitive processes that test-takers engage in during their responses (AERA, APA, NCME, 1999). Similarly, if the directions of the test are either not clear, or are not adhered to with a high degree of compliance, then the test results might not accurately reflect the true variance among individuals (Furr & Bacharach, 2013). Thus, evidence should be established that respondents take the test or survey in a manner consistent with how it is intended, and that the cognitive processes that direct their responses are consistent. Several authors have argued that conducting cognitive interviews with participants from the population of interest are an informative way to address the issue of differing response processes (AERA, APA, NCME, 1999; Arjoon, Xu, & Lewis; 2013; Desimone & Le Floch, 2004).

Evidence for relationships with other variables. A psychological attribute or process does not exist alone, but rather operates in concert with other processes. Hence, there are relationships among various psychological attributes – some that are positive and others that are negative. For instance, depression is more likely to be correlated with anxiety than happiness will be. It could be expected, then, that scores from a measure of depression would be positively correlated with scores from a measure of anxiety. When two theoretically related constructs (or the same construct from two different tests) are found to empirically correlate with each other – that is referred to as convergent validity evidence (Furr & Bacharach, 2013; Murphy & Davidshofer, 2005). Conversely, discriminant validity evidence is when two unrelated constructs are empirically found to not correlate with each other (Furr & Bacharach, 2013). For example, it would be expected that general self-esteem and attitude toward science would not correlate to a high degree. This type of evidence is very important when developing a survey, because part of ensuring that a specific attribute is being measured includes being able to demonstrate viable relationships with theoretically related variables and discriminate between theoretically unrelated variables.

Evidence for criterion-related validity. Similar to evaluating the relationships with other variables, test scores may also be evaluated for their relationships to external criteria. Criteria in this case refer to “a measure that could be used to determine the accuracy of a decision” (e.g., success in college, job performance) (Murphy & Davidshofer, 2005, p. 180). Two sub-types of

validity evidence have been described in the literature: evidence of predictive validity and evidence of concurrent validity (Martella et al., 2013; Murphy & Davidshofer, 2005). Strategies for assessing evidence of predictive and concurrent validity are nearly the same, with a few differences. The most distinguishing factor between the two types of strategies is that predictive validity strategies utilize scores from a random sample in a given population (i.e., applicants to college, job seekers), and concurrent validity strategies utilize scores from an already intact sample (i.e., students in a college class). A second difference between the two is that decisions are made without test scores when assessing predictive validity, and decisions are typically made at the same time with concurrent validity strategies (Guion & Cranny, 1982). To illustrate the difference, consider the following example.

Suppose a test is designed in order to predict success in college. If a test is designed to predict success in college, then the validity of scores from the test depend on the positive relationship between test scores and success in college. To assess evidence of predictive validity, the researchers would need to accept students to the college without examining the scores from the test. The test could be administered either prior to or following acceptance to the college. In this way, the test scores are likely to be more varied because students would be accepted regardless of their scores, instead of accepting students based on their high scores (concurrent validity evidence). Thus, the correlation between success in college and the test scores would be more accurate for the general population of applicants using predictive validity strategies as opposed to concurrent validity

strategies (Murphy & Davidshofer, 2005). There are, however, some practical constraints to the predictive validity approach. There would be a risk associated with employers and colleges accepting applicants without some measure of their potential success in either arena. Hence, in most cases, concurrent validity strategies are employed over predictive validity strategies (Murphy & Davidshofer, 2005).

While criterion-related validity evidence is accepted by most, some authors warn against the tempting exclusive use of this form of validity evidence in the development and evaluation of their tests (Furr & Bacharach, 2013; Messick, 1989). Furr and Bacharach (2013) argue that some developers might write a test that is only intended to predict job performance. If found to be predictive of job performance, then the test is called a “valid” measure. By doing so, the authors argue, the test developers are ignoring the underlying psychological construct and are not concerned with what the scores from the test really mean. With such a narrow approach, the interpretations from the measure cannot be considered valid by following the contemporary practices for assessing validity evidence (AERA, APA, NCME, 1999). However, criterion-related validity evidence centers on an important implication of giving tests and the resulting test scores – the decisions made because of them (Murphy & Davidshofer, 2005). Thus, as part of an overall evaluation of validity evidence, criterion-related evidence has a valuable place, especially for tests that are used in decision-making.

Evidence for internal structure validity. The internal structure of a test or survey has to do with the intercorrelations of the items, or how well the items cluster together. Some psychological constructs contain multiple dimensions, and a well-written test will reflect those different dimensions. The extent to which the internal structure of a test represents the dimensionality of the construct being measured is an indication of the degree of evidence for internal structural validity (Furr & Bacharach, 2013; Messick, 1989). Once a test has been assigned the appropriate number of dimensions, as outlined by the theoretical construct, evidence should be presented to empirically confirm the internal structure of the test (AERA, APA, NCME, 1999). Usually this is accomplished with factor analysis methods, as discussed above. However, the important implication for validity here is not just that the empirical data confirm or repudiate the hypothesized structure of the test, but rather that the test scores are interpreted in a way that is consistent with the structure of the construct domain (Loevinger, 1957; Messick, 1995). For instance, if a survey of self-esteem were found to contain four separate factors, consistent with theory, then the scores should be interpreted as four different values. A total score for a multidimensional test is usually not appropriate, unless the underlying theory states otherwise.

Summary. The amount of evidence for validity should not be considered a threshold to be reached or a finish line to be crossed. Instead, the evidence should be considered as strong and weak, with a relative assessment as to what determines strong enough evidence for the particular application of measurement (Furr & Bacharach, 2013). At the core of test validity is construct validity, defined

as the interpretations of scores from tests obtained under a specific set of conditions (AERA, APA, NCME, 1999). This should always be considered with paramount importance when carrying out a study where measurement is involved or when developing a measurement tool. In addition, when considering what constitutes “strong enough” evidence for validity, the researcher should consider the decisions that will be made as a result of test scores. Certainly, a test that will dictate the future of a prospective college student should demonstrate greater evidence of validity for the construct being measured than a classroom survey designed to inform the instructor of the level of interest in geology that exists among his or her students. Regardless of the application or strength of validity, the researcher is implored to provide sufficient evidence to support any claims made related to the interpretations and scoring associated with use of the a given test (AERA, APA, NCME, 1999).

Reliability

Reliability of a test can refer to several distinct, but related aspects: (1) the consistency of scores across repeated administrations, (2) the degree to which items measure the same construct, (3) the correlation between a set of observed scores and the true scores, or (4) the correlation of scores between analogous test forms of the same construct (Martella et al., 2013). While seemingly different, all of these aspects of reliability are rooted in the same fundamental concept that reliable scores are those that have minimal measurement error. Thus, the purpose of assessing reliability is to estimate the error of measurement (Murphy & Davidshofer, 2005). Reliability is not a property of tests, as is sometimes

erroneously reported, but rather a property of scores (Vacha-Haase, Kogan, & Thompson, 2000). Hence, it is recommended that reliability estimates be calculated for every new sample that takes the test, especially if they differ from the population for which the test was intended (AERA, APA, NCME, 1999; Vacha-Haase et al., 2000). Furthermore, for tests that are multidimensional, each subscale represents a different construct. Thus, the reliability estimate should be calculated for each subscale of a test (AERA, APA, NCME, 1999).

Reliability theory is predicated on the hypotheses that the true value of a measurement can never be known and that error associated with the measurement is random (Furr & Bacharach, 2013; Henson, 2001). Thus, reliabilities are not actual values, but rather, estimates of actual values. By definition, a “true score” is the average score on a test that has been administered to an individual an infinite number of independent times (AERA, APA, NCME, 1999; Henson, 2001). The general equation for calculating reliability is: observed score = true score + random measurement error.

Random measurement error, rather than systematic error, negatively affects the reliability estimates, because systematic error would cause scores to vary consistently among the test-takers (Furr & Bacharach, 2013). For instance, a test scorer that erroneously left out one item from the composite score of all test-takers would cause invalidity in the scores, but this accidental act would not affect the reliability of the scores. However, in cases where the reliability is negatively affected, the validity of the interpretations from scores is always called into question. Unreliable scores cannot produce valid interpretations, because it

is impossible to measure true variance of a psychological attribute with high levels of random measurement error (AERA, APA, NCME, 1999; Furr & Bacharach, 2013; Martella et al., 2013).

Several forms of reliability are presented in the literature: coefficient of internal consistency, coefficient of stability (test-retest reliability), and coefficient of equivalence (parallel forms reliability). It has been suggested that the most popular form in psychological research is the coefficient of internal consistency because only one test administration is necessary (Henson, 2001). The other two types require at least two independent test administrations.

The coefficient of stability and coefficient of equivalence reliability estimates are calculated at the test-level, meaning that the entire score is taken into account when generating these estimates. This is possible because at least two independent total scores for each individual are required. The coefficient of equivalence reliability is when two parallel tests are given, and the correlation between the scores on both tests is calculated. The calculated correlation is the estimate of reliability using this method. An inherent problem exists, however, in that one cannot ever know if the tests are truly parallel. Furr and Bacharach (2013) state, “two tests are considered parallel only if (a) they are measuring the same set of true scores and (b) they have the same amount of error variance” (p. 126). In short, this means that the tests must measure the same attribute and they both must be designed in such a way as to not introduce more error variance in one test versus the other. Although the coefficient of equivalence can be valuable when administering the tests at very short time intervals, the

coefficient of stability avoids some of the pitfalls inherent in the former method. The coefficient of stability (often referred to as test-retest reliability estimate) is estimated when the same test is given to a sample of respondents twice on different occasions, and the correlation between their scores is calculated. As with the coefficient of equivalence, the correlation between the two scores is the reliability estimate. Typically, an appropriate time frame between administrations is 2 to 8 weeks (Furr & Bacharach, 2013). However, this recommended time frame is dependent on the construct being measured. For example, a test measuring content knowledge may require a shorter time interval between testing, as the researcher would want to minimize new knowledge acquisition. The test-retest method avoids the problem of different tests potentially measuring different constructs, but still has its own drawbacks. The most problematic is the instability of true scores. For example, a student may differ in her level of depression between the three weeks that separate the two test administrations. This reflects a difference in her true score from one time to the next. However, the method of estimation cannot distinguish between true score differences and random measurement error. Thus, the test-retest estimate will be low, which could be an inaccurate estimate of the actual reliability of scores produced by the test (Murphy & Davidshofer, 2005). Thus, it is recommended to interpret test-retest reliability estimates with caution, especially when measuring highly fluctuating psychological attributes (Furr & Bacharach, 2013).

Internal consistency is related to the homogeneity of the items, or how well they describe a given construct (Henson, 2001; Martella et al., 2013). Estimates

for internal consistency include the split-half estimate, Küder-Richardson formulas, and Cronbach's alpha (α). The split-half estimate involves splitting the items in half (usually even-odd) and computing the correlation between the scores of each group. Afterwards, the correlation is placed into the Spearman-Brown formula, which then gives an estimate of the reliability (Furr & Bacharach, 2013). This formula is needed because split-half estimates are at the "half test-level, meaning that the correlation is only representative of half of the test. If the items and corresponding scores in both groups are similar in nature, then the split-half estimate should be high between scores of the different items. The Küder-Richardson (KR) formulas and Cronbach's alpha are both internal consistency estimates at the item-level. They have been described as averages of all possibilities of split-half estimates (Martella et al., 2013). The KR formulas are used with tests having binary item scores, and Cronbach's alpha is used with non-binary item scores. To calculate the Cronbach's alpha involves a two-step process whereby the covariances of each item pair are calculated then summed together, then placed in a formula with the variance of scores from the entire test (Furr & Bacharach, 2013). The KR formulas are calculated in a slightly different way that reflects the special character of binary scores (Furr & Bacharach, 2013). Both KR formulas and Cronbach's alpha are heavily used estimates of internal consistency in the social sciences.

The importance of reliability in measurement cannot be overstressed. As stated earlier, decrements in reliability will negatively affect and limit the interpretations that can be drawn from measurement data. As a further

consequence, the differences between perfectly unreliable scores can never be statistically significant, no matter how large the sample size (Reinhardt, 1996). Although the true reliability of test scores can never be known, the methods presented here have been widely used in estimating the true reliability and are theoretically sound (Furr & Bacharach, 2013).

Existing Motivation Instruments and Scales

Motivation, as a whole, is made up of many different processes and constructs. Authors have differing theoretical perspectives about how these factors add up together to compel an individual to behave in a certain way (Ryan & Deci, 2000b; Schunk, 2008; Schunk & Pajares, 2002; Wigfield & Eccles, 2002a). Whichever viewpoint is endorsed, the individual constructs must be measured in a way that leads to valid and reliable conclusions. This section of the review will highlight several instruments and scales that have been developed to measure self-efficacy, interest, and effort beliefs. Due to the overwhelming number of scales and instruments that have been used in studies of academic motivation, only those which are most relevant to this dissertation study will be reviewed.

Motivated Strategies and Learning Questionnaire. Perhaps one of the most renowned instruments of motivation research is the Motivated Strategies and Learning Questionnaire (MSLQ). Originally developed by Pintrich and colleagues (1993), the (MSLQ) is comprised of 81 items across six motivation subscales and nine learning strategies subscales. It has been used in well over 50 studies and translated in a number of languages (see Duncan and

McKeachie, 2005 for review). The items are measured on a seven-point Likert-type scale from 1 (not true at all of me) to 7 (very true of me). Confirmatory factor analysis was used to quantitatively assess whether the items from a particular subscale belonged together as one factor. This technique is associated with supporting the validity of scores from an instrument or scale. A sample of 356 undergraduates from many disciplines volunteered to take the survey following its initial development. All of the items were demonstrated to adequately “fit” with the corresponding subscale in the initial published study. Of the motivation subscales, one is related to self-efficacy – *self-efficacy for learning and performance*, and another is related to interest – *task value*. Example items from the self-efficacy scale are, “I expect to do well in this class”, and “I’m confident I can do an excellent job on the assignments and tests in this course”. Example items from the task value scale read, “It is very important for me to learn the material in this course”, and “I like the subject matter of this course”. Both the *self-efficacy for learning and performance* and *task-value* subscales were shown to generate scores with an excellent Cronbach’s alpha estimate ($\alpha \geq 0.90$). Overall, the MSLQ has a reputation as an exceptional instrument to measure motivation and learning strategies across many distinct constructs.

Chemistry Attitudes and Experiences Questionnaire. The Chemistry Attitudes and Experiences Questionnaire (CAEQ) is composed 69 items across three major scales: *self-efficacy* (17 items), *learning experiences* (31 items), and *attitude toward chemistry* (21 items) (Dalgety & Salter, 2002). The *learning experiences* and *attitude toward chemistry* both contain numerous subscales.

Two subscales in the *attitude toward chemistry* scale, *leisure interest* and *career interest*, both relate to personal interest. Example items from the *self-efficacy* scale include reporting confidence levels of “Tutoring another student in a first year chemistry course”, or “Choosing an appropriate formula to solve a chemistry problem”. An example item from the *leisure interest* subscale is, “The tutorial sheets helped me understand the lecture course.” An example item from the *career interest* subscale is, “The material presented in tutorials was useful.” The CAEQ was specifically developed for use in an introductory college chemistry setting. Exploratory factor analysis was employed to evaluate the construct validity of data gathered from an initial sample of participants (Dalgety et al., 2003). The results revealed that most items loaded to the appropriate factor, but a few items correlated with multiple factors. Furthermore, the authors interviewed 19 students to obtain feedback on the readability and comprehension of the items, a technique used to establish response process validity (AERA, APA, NCME, 1999). However, no data from these interviews was provided in the publication. The authors also tested the differences in scores between students who were planning on enrolling in second year chemistry versus those who were not. It could be expected that those students planning to enroll in a second year of chemistry would have higher self-efficacy and more positive attitudes toward the course. Indeed, on every subscale of the instrument, those planning to enroll in a second year of chemistry scored significantly higher than those who were not. The CAEQ represents one of just a few instruments that measures constructs related to motivation, which was specifically designed for chemistry.

Despite this, only a few published studies have used the instrument or its subscales (Dalgety & Coll, 2006a, 2006b; Villafane et al., 2014; Winkelmann et al., 2015).

Colorado Learning Attitudes about Science Survey. Developed originally for use in college-level physics (Adams et al., 2006), the Colorado Learning Attitudes about Science Survey (CLASS-Phys) has also been adapted for use in college chemistry (CLASS-Chem) (Barbera et al., 2008) and biology (CLASS-Bio) (Semsar, Knight, Birol, & Smith, 2011). During the original development process, experts and over 40 students were interviewed to generate the 42 item CLASS-Phys. Following the interviews, factor analysis was performed to confirm that the appropriate items related adequately to their respective category. The CLASS-Chem added an additional 8 items, bringing the total to 50 items across 9 subcategories. Interviews were conducted with student volunteers to establish consistent readability and interpretation across all items. The resulting subcategories cover several areas of students' beliefs in chemistry including: personal interest, problem solving, conceptual learning, and atomic-molecular perspective of chemistry. Example items from the personal interest subcategory are, "I think about the chemistry I experience in everyday life" and "I am not satisfied until I understand why something works the way it does". The CLASS-Chem was given to over 50 chemistry faculty members at many different institutions to provide feedback, as well as to establish the "expert response". Factor analysis was performed in the same manner as with the CLASS-Phys. Later, the CLASS-Chem was subjected to a more thorough psychometric

evaluation (Heredia & Lewis, 2012). Heredia and Lewis (2012) argued that the scores from the nine scales could not be interpreted as individual scores unless the scales are administered separately. The CLASS-Chem items have been used in a few studies since their adaptation (Phillips & Grose-Fifer, 2011; Schaller et al., 2015; Winkelmann et al., 2015)

Interest scales. The interest scales reviewed here are not part of a named instrument, but the items from these scales (or adaptations of these items) have been used in multiple studies (Harackiewicz et al., 2008; Kim & Schallert, 2014; Linnenbrink-Garcia et al., 2010; Linnenbrink-Garcia, Patall, & Messersmith, 2013; Plass et al., 2013) and were developed by experts of interest research. Although several interest scales are reported in these articles, only the maintained interest and initial interest scales will be discussed. The initial interest scale, originally published by Harackiewicz et al., (2008), contained seven items and was measured on a 7-point Likert-type scale ranging from 1 (not at all true of me) to 7 (very true of me). The scale was given to a sample of over 850 undergraduate psychology students. The only psychometric property reported for the scores associated with this scale was Cronbach's alpha ($\alpha = 0.90$), as an indicator of internal consistency. Example items from this scale include, "I chose to take Introductory Psychology because I'm really interested in the topic", and "I think the field of psychology is an important discipline." The maintained interest scale, referred to as "hold" in Harackiewicz et al. (2008) consisted of nine items on the same Likert-type scale as the initial interest scale. As part of the same study (Harackiewicz et al., 2008), but in a separate publication, confirmatory

factor analysis was applied to test whether the items were best described as one factor (maintained interest) or two distinct factors (maintained interest-feeling and maintained interest-value) (Linnenbrink-Garcia et al., 2010). The authors determined that items that describe the value component of maintained interest (“I think that what we are learning in this course is important”) were quantitatively distinct from those describing the feeling component (“I think the field of psychology is very interesting”). This model is consistent with Schiefele’s (1991) conception of interest as containing both feeling-related and value-related valences. The items from both the initial and maintained interest scales reflect the commonly accepted modern conceptions of individual interest (Renninger, 1992) and situational interest (Krapp, 2002).

Science Motivation Questionnaire. The Science Motivation Questionnaire (SMQ) was developed by Glynn and Koballa (2006). The SMQ contains 30 items evenly distributed across six scales: *intrinsically motivated science learning, extrinsically motivated science learning, relevance of learning science to personal goals, responsibility, confidence, and anxiety about science assessment*. In the initial publication (Glynn & Koballa, 2006), the authors only report internal consistency for the instrument ($\alpha = 0.93$). However, since there are six scales, this is not appropriate, as internal consistency needs to be assessed for all scales individually. In a later study, Glynn et al., (2009) conducted an exploratory factor analysis (EFA) on the SMQ using data gathered from over 750 nonscience majors. As a result of the EFA, student interviews, and essays written by students, the six factors of the questionnaire were trimmed to

five with an addition and two combinations – *intrinsic motivation and personal relevance, self-efficacy and assessment anxiety, self-determination, career motivation, and grade motivation*. Following further revisions with a focus group of students and a small pilot test sample, 14 items were removed from the original SMQ and 9 additional items were added. The resulting instrument was the 25-item SMQ II with five, 5-item scales: *intrinsic motivation, career motivation, self-determination, self-efficacy, and grade motivation*. Exploratory and confirmatory factor analyses were conducted with results indicating the data fit the models well. Example items from the *self-efficacy* scale include: “I believe I can earn a grade of “A” in science”, and “I am sure I can understand science”. Some of the items from the *intrinsic motivation* scale relate to interest such as, “Learning science is interesting”, and “The science I learn is relevant to my life”. The SMQ II has been used in a few studies since the revisions (Psycharis, 2013; Salta & Koulougliotis, 2015), adapted to chemistry (Tosun, 2013), and one study used the original SMQ (Zeyer et al., 2013).

College Chemistry Self-Efficacy Scale. The College Chemistry Self-Efficacy Scale (CCSS) was originally written in English, but has only been administered in Turkey. Published in 2009, the CCSS represents the first effort to develop a stand-alone self-efficacy scale for use in college chemistry (Uzuntiryaki & Aydin, 2009). The CCSS consists of 21 items across three sub-scales: *self-efficacy for cognitive skills* (12 items), *self-efficacy for psychomotor skills* (5 items), and *self-efficacy for everyday applications* (4 items). The nine-point rating scale was Likert-type ranging from 1 (very poorly) to 9 (very well). The

developing authors performed both exploratory and confirmatory factor analyses with independent samples of introductory chemistry students to assess the internal structure of the scale. The data were found to fit the models well and confirmed that the items from each subscale correlated well among each other. In addition, the authors demonstrated that chemistry majors scored higher on the self-efficacy scale than non-majors, a result that is expected. Example items from the *self-efficacy for cognitive skills* sub-scale include: “To what extent can you explain chemical laws and theories?” and “How well can you choose an appropriate formula to solve a chemistry problem?” An example item from the *self-efficacy for psychomotor skills* sub-scale is, “How well can you work with chemicals?” Finally, an example item from the *self-efficacy for everyday applications* is, “How well can you recognize the careers related to chemistry?” Several studies have reported using the CCSS in a Turkish college setting (Aydin et al., 2011; Uzuntiryaki, 2008; Uzuntiryaki & Capa-Aydin, 2013). The CCSS was thoughtfully developed, with underlying theory in mind (Bandura, 1997; Pajares, 1996), and is a valuable tool for assessing self-efficacy in introductory college chemistry courses.

Effort beliefs scale. The effort beliefs scale, like the interest scales described above, is not part of a larger instrument. Rather, the nine items were written by experts of implicit theories of intelligence (Sorich & Dweck, 1997) and used in several subsequent studies, either in the original wording (Blackwell, 2002; Blackwell et al., 2007; Tempelaar et al., 2015), or a subject-specific adaptation (Jones et al., 2012). The rating scale is a 6-point Likert-type, ranging

from 1 (disagree strongly) to 6 (agree strongly) (Blackwell et al., 2007). There are no data indicating a psychometric evaluation for these items other than reliability estimates (Cronbach's alpha) reported in each of the studies. Four items were written positively ("If you don't work hard and put in a lot of effort, you probably won't do well"), and five items were written negatively ("It doesn't matter how hard you work – if you're not smart, you won't do well"). This scale represents only one of two effort beliefs scales found to be published in the literature. The other scale, published in a study by Stipek and Gralinski (1996), is more geared toward lower age groups with items such as, "Anyone who works hard could be one of the smartest in the class". Although this effort beliefs scale (Sorich & Dweck, 1997) was written for use with 7th grade students, it has been successfully used at the high school (Jones et al., 2012), and college level (Tempelaar et al., 2015).

Summary

Motivation and its processes represent a complex, multifaceted fabric of interwoven psychological constructs. As such, there is tremendous connectivity among competing theories that attempt to explain human behavior in terms of their underlying motives. There are many reasons and explanations as to why someone would engage with, or avoid a given task or domain. Self-efficacy, interest, and effort beliefs are three distinct constructs, but are connected by the collective impact they can exert on one's decision-making and behavior. Consider the following scenario to illustrate how self-efficacy, interest, and effort

beliefs can be tied together in a student's psyche as influential motivational processes.

Jane has been interested in chemistry since she was 12 years old when she was given a home chemistry lab kit. She decided to pursue chemistry in high school and did well while enjoying it. At the start of her freshman year in college, Jane was confident she would do well, and driven to work hard. However, she suffered a loss of confidence in her ability as a result of a failing grade on the first exam. Despite this, she continued to work hard studying for her class, and performed better on the next two exams. This was due in part to the class structure that allowed her to ask questions in class during group work sessions. By the final exam, Jane was more confident than ever in her ability to successfully complete the problems assigned in her chemistry course. As a result, she finished the course with a B average.

This example represents someone whose individual interest brought her to chemistry, her effort beliefs carried her through a trying period; and as a result of her hard work, she became more confident and expectant of success by the end of the course than when she started. Although this seems like a simple deconstruction, in reality, all of these processes were occurring simultaneously, along with myriad other psychological and physiological events.

It is the complexity of motivation that makes it such a rich and daunting research pursuit. In spite of this, much can be learned and improved upon by the study of motivation, particularly domain-specific motivation. Students will conduct

themselves in a manner that is a reflection of their motivational state. By changing the way courses are taught, in a way that enhances students' motivation in chemistry, perhaps the success rate would increase, both in the quality and quantity of students that emerge from a program. Whether it is a complete overhaul to the teaching practice, such as POGIL, or small changes to the way feedback is given, or even just spending half of a class period on discussing how intelligence can grow with effort, all of these could affect students in more ways than just their performance on exams. However, before any changes can be shown to be effective toward enhancing self-efficacy, for example, there must be appropriate tools in place to measure self-efficacy in chemistry. The evidence must extend beyond theoretical and hypothetical rhetoric. The focus of this dissertation study will be to establish evidence for such tools, so that motivational constructs can be accurately and reliably measured in the context of introductory chemistry.

CHAPTER III

METHODOLOGY

Introduction

In order to appropriately answer the research questions guiding this study, a variety of methods were employed, both qualitative and quantitative. The quantitative data were comprised of first-semester general chemistry students' responses to initial and revised scales linked to academic motivation, as well as course performance data. The qualitative data were comprised of interviews with students from the same target population. The quantitative methods were used to draw inferences from patterns in students' scores as well as to formulate models that describe the internal structure of measurement scales related to motivation. The qualitative threads are important for enhancing the meaning and credibility of the quantitative results, as well as for informing revisions to the quantitative models. The study consists of three major phases, ordered chronologically. The first phase of the study involved testing the unabbreviated scales with samples from the target population in order to establish evidence of validity and reliability for data from the scales. The second phase of the study involved using the scales, with any associated modifications, in a classroom-based study where a few interventions will be implemented. The final phase of the study was almost

entirely qualitative and aimed at enriching the current understanding of chemistry students' effort beliefs.

Research Questions

The major research goals of the present study were to adapt a set of brief scales designed to measure students' self-efficacy, effort beliefs, and interest for use in general chemistry, and then use these scales in a classroom-based investigation with interventions. There are seven research questions that are addressed by the current study:

- Q1 What modifications are needed to produce brief, chemistry-specific scales of self-efficacy, interest, and effort beliefs?
- Q2 What evidence supports the functioning of each of the modified scales?
- Q3 To what extent do students' self-efficacy, interest, and effort beliefs change across the first semester of general chemistry?
- Q4 To what extent are students' self-efficacy, interest, and effort beliefs affected by brief interventions targeting their values and implicit theories of intelligence?
- Q5 What are the connections among self-efficacy, interest, and effort beliefs with general chemistry students?
- Q6 To what extent do self-efficacy, interest, and/or effort beliefs predict course performance in general chemistry?
- Q7 What are the sources and influences of effort beliefs toward chemistry among general chemistry students?

Protection of Human Subjects

In accordance with the Institutional Review Board (IRB) at the University of Northern Colorado, approval was requested for any data collection involving human subjects. The appropriate applications were submitted to the IRB prior to

collecting any data (see Appendix A), and data collection did not commence until approval was received from the IRB. Following approval, the voluntary nature of the data collection was explicitly stated and emphasized to all potential participants during invitations by the researcher (see Appendix B). Once collected, all sensitive data with personal identifiers was stored on a password-protected computer, whereby only the researcher had access. Any non-digital data with personal identifiers was stored in locked cabinets, such that only the researcher had access. Any data that was made available to other researchers was formatted such that the personal identifiers were replaced with numerical identifiers. Any interview transcripts were labeled with random numbers to protect the identity of the participants. Throughout all phases of this dissertation study, confidentiality was maintained for all participants and the voluntary basis for collecting data were emphasized.

Phase One: Adaptation of Scales for Use in General Chemistry

Research Design

The research design employed in this phase of the study was that of mixed-methods. Mixed-methods research is a blending of quantitative and qualitative methods, such that the results from both will be more informative than either approach alone. The intent of mixed-methods research is not just to collect data in both a qualitative and quantitative fashion, but also to integrate the findings as a means to strengthen the overall research narrative (Creswell, 2009). Six major mixed-methods designs have been described in the literature: sequential explanatory, sequential exploratory, sequential transformative,

concurrent triangulation, concurrent nested, and concurrent transformative (Creswell, Clark, Gutmann, & Hanson, 2003). The specific design employed in this phase of the study was concurrent nested, whereby qualitative and quantitative data were collected and analyzed at the same time. Usually, in the framework of a concurrent nested design, one form of data will take priority over the other. In this phase of the study, quantitative data were the primary source of evidence for model testing, and the qualitative findings were more supportive in nature. This type of design certainly did not diminish the importance of qualitative data. Instead, the quantitative strand was set up as the foundation for the model and the qualitative strand informed ways the model was revised to better encompass the meaning of the theoretical constructs. The goal of model revision was to have the most parsimonious model, whereby all unnecessary and less meaningful items have been removed. To accomplish this, interviews with students and scale responses had to be considered together.

Scales

The scales that were used in this study are based on constructs best operationalized in a context-specific manner. The scales were taken from previously published studies in a variety of disciplines. All items, except those from the self-efficacy scale, were modified to reflect a chemistry-specific context. The self-efficacy items were excluded because they were originally developed for use in college chemistry (Uzuntiryaki & Aydin, 2009).

Preliminary wording changes. The first step to adapting the scales for use in general chemistry involved making minor wording changes to each item.

For example, the interest scale was developed for use in college psychology classes, and the effort beliefs scale was developed for high school math. Hence, the items reflect that by including references to the subjects of psychology and math. For most items in the interest and effort beliefs scales, the only changes necessary were to replace the words “psychology” or “math” with the words “chemistry” or “general chemistry”.

Chemistry self-efficacy scale. The self-efficacy scale was taken from the College Chemistry Self-Efficacy Scale (CCSS) (Uzuntiryaki & Aydin, 2009). These items are designed to measure a student’s judgment of his or her ability to complete a given task in a chemistry course. The original instrument (21 items) has three subscales of chemistry self-efficacy: self-efficacy for cognitive skills (12 items), self-efficacy for psychomotor skills (5 items), and self-efficacy for everyday application (4 items). The original items are on a 9-point Likert-type scale ranging from “very poorly” to “very well.” Select items from the self-efficacy for cognitive skills subscale were evaluated for this study. These items are intended to measure students’ belief in their ability to perform intellectual operations in chemistry (Uzuntiryaki & Aydin, 2009). The main focus of this study was related to chemistry problems encountered during the lecture portion of class; thus, items related to the laboratory or the nature of science were excluded. An example item that was excluded is: “How well can you write a lab report summarizing main findings?” Example items that were retained include: “To what extent can you explain chemical laws and theories?” and “How well can you read the formulas of elements and compounds?” In the original instrument

developed by Uzuntiryaki & Aydin (2009), nine numerical choices were given, but only five delineated categorical choices, ranging from “very poorly” to “very well”, were placed above the numbers. The student then is left with multiple numerical choices per category. The meaning of the difference between the two numerical choices is therefore lost. It stands to reason that if only five categories are given, then only five numerical choices are necessary. As there was no compelling reason to retain the nine options, and to allow for electronic scoring, the nine-point Likert-type scale used in the original instrument was adjusted to a five-point Likert-type scale. Since the internal structure of the scale was tested, a Confirmatory Factor Analysis (CFA) result consistent with the CFA from the original authors (Uzuntiryaki & Aydin, 2009) was sufficient to provide favorable support for the condensed number of response options. In addition, there is evidence that changing scale length does little to affect the distribution about the mean, skewness, or kurtosis (Dawes, 2008).

Initial and maintained interest scales. The original initial interest and maintained interest scales were developed by Harackiewicz et al. (2008) for use in college psychology. The initial interest items are designed to measure a student’s interest in psychology at the beginning of an introductory undergraduate psychology course. The maintained interest items were given to students at week 13 of the semester and were designed to measure the “hold” component of situational interest. Exploratory and confirmatory factor analysis performed with these items confirmed the distinction between the “catch” and “hold” components of situational interest (Linnenbrink-Garcia et al. 2010). The

original scales for initial interest and maintained interest have seven and nine items, respectively. Both original scales are measured on a seven-point Likert-type scale ranging from “not at all true of me” to “very true of me.” The wording of these items was modified slightly to fit the context of a chemistry course, mostly by just replacing the word “psychology” with the word “chemistry.” As the authors provided no rationale for retaining the 7-point scale, and to keep the number of response options consistent across all measures, the scale was adjusted to 5-points. Lastly, the responses on the scale were changed from “true of me” statements to “agree” statements as agree-type response options better fit the wording of the items.

Effort beliefs scale. The original items for the effort beliefs scale were developed by Sorich and Dweck (1997) and first used in Blackwell’s (2002) unpublished doctoral dissertation study, which involved seventh grade students. The nine-item effort beliefs scale was designed to measure the degree to which students believe their effort will lead to positive outcomes. These items were then adapted for use in a motivational study involving ninth grade math students (Jones et al., 2012). The effort beliefs scale used by Jones et al. (2012) consisted of nine items measured by a six-point Likert-type scale ranging from “strongly disagree” to “strongly agree.” The exact wording of each item was used (based on Jones et al. (2012) version), except for substituting the word “chemistry” for the word “math.” The scale consisted of four positive items (“If a chemistry assignment is hard, I’ll probably learn a lot doing it”), and five negative items (“To tell the truth, when I work hard at chemistry, it makes me feel like I’m

not very smart"). In addition, the scale range was adjusted from a 6-point to a 5-point Likert scale to remain consistent with the other scales.

Participants

The target population for the adapted scales was first-semester general chemistry students. Due to logistical and resource-based constraints, convenience sampling was used. Hence, the sample was not representative of general chemistry students everywhere (Crotty, 1998). However, student samples were gathered from more than one institution and during multiple semesters. Students were solicited for participation from both mid-sized and large universities. Students from the mid-sized university were sampled in both the initial and cross-validation data collections, whereas those at the large university were only sampled in the cross-validation data collection. Although students were sampled from the same course throughout the study, sampling did take place in different settings (either lecture or lab) based upon permission granted to the researcher. Hence, it is important to point out that students who were enrolled in lecture and not lab or vice versa might not have been included in the study if the sampling occurred in the setting where they were not enrolled.

Initial Survey Data Collection

During phase one of the study, participants were recruited from the laboratory sections of first-semester general chemistry at a mid-sized university located in the Rocky Mountain west. The data collection took place during lab at two time points: week 1 and week 13. Teaching assistants (TAs) in each section were given a statement from the researcher explaining the purpose of the

research project as well as a notice of confidentiality. TAs were asked to read this statement to their students prior to distributing the scales in their sections. Afterwards, students were asked to participate in the study, which required them to complete a series of scale items regarding their motivational beliefs and feelings about chemistry. Students were handed a consent form with the scales and told that their responses would only be included in the study if they provided consent. A student identifier was requested from participants, so that responses from time 1 (week 1) could be tied to responses from time 2 (week 13). Demographic items were included in the packet, as well as an item asking students if they would like to participate in an interview regarding the scales. Students were given approximately 15 minutes to complete the scale items at both time points. For students to be included in pre- and post-semester comparisons, they must have been present in lab for both data collections. Students who were not enrolled in a laboratory section, or who were not present on the first day of lab were excluded from the study.

Cross-Validation Survey Data Collection

The initial data collected was used to test models based on the internal structure of each scale. Because items were dropped such that the scale better fit the proposed model, the models were cross-validated with an independent sample. Cross-validation studies are recommended anytime a model has been re-specified (Brown, 2006). If a sample size is large enough, it could be split into two independent samples, one to justify any changes to the model, and the other to cross-validate the re-specified model (Brown, 2006). However, the sample size

for the initial data collection was not large enough to split. Consequently, a second sample of participants was recruited during a subsequent semester following the initial data collection.

Survey Data Analysis

Descriptives. Descriptive statistics were analyzed on all data to determine means and standard deviations as well as to check for skew and kurtosis. Descriptive statistics are important to analyze prior to any inferential statistics, as they provide the researcher with a look at the central tendency of the data. Non-normal data with excessive skew and kurtosis cannot be analyzed in the same way as normal data (Martella et al., 2013). In line with what is commonly recommended, acceptable skew and kurtosis values were considered as falling in the range of -1 to +1 (Huck, 2012). Reliability estimates for internal consistency (Cronbach's α) were also calculated for each scale as well. Cronbach's α is an estimate of the internal consistency in the responses and should be reported with respect to each scale (AERA, APA, NCME, 1999). A value of 0.70 is considered acceptable for classroom multiple-choice tests and rating scales (Murphy & Davidshofer, 2005). Statistical Package for the Social Sciences (SPSS) 20.0 software was used for these analyses.

Time 1 and time 2 measurements. Data were gathered at two time points (week 1 and week 13) from the initial sample and a subset of a separate cross-validation sample. Only students who completed the scales at both time points during the initial and cross-validation data collections were able to be included in these analyses. All item scores from each scale were aggregated to

produce a scale score. The mean score aggregation method was used as each scale contains a different number of items. Analysis of variance (ANOVA) tests were performed on each scale at both time points separately to determine if any significant differences existed among students' scores based on major choice and gender. Majors were grouped into categories (chemistry, other science, non-science, and other). Before evaluating the ANOVA tests, the assumptions of normality and homogeneity of variance were checked. A normal probability plot of the residuals was used to check for normality, and Levene's test was used to check for homogeneity of variance. Both assumptions were met, so the ANOVA tests were examined for significance. Those ANOVA tests that were significant were followed up with Tukey's pairwise *post-hoc* tests. The ANOVA tests indicated if there is a significant difference in a given dependent variable (self-efficacy, interest, effort beliefs, performance) among any of the groups tested. The Tukey's pairwise tests indicated which specific groups differed significantly with respect to the dependent variable. Tukey's *post-hoc* test was used because Tukey's test is very common and is the preferred test for pair-wise comparisons with one-way ANOVAs (Gray & Kinnear, 2012). Paired samples *t*-tests were conducted with scores from all scales to determine if any changes were significant in students' scores across the semester. To further assess if these changes were different by major or by gender, mixed-between-within ANOVA tests were employed for the interest and self-efficacy scales, followed by *t*-tests for each group. All tests were evaluated at $p < 0.05$.

Confirmatory factor analysis. Confirmatory factor analysis (CFA) is a powerful tool for assessing how well a proposed model fits a set of measured variables. Each scale (i.e., self-efficacy, initial interest, and effort beliefs) was considered a latent variable (or factor) and each item, an indicator for its respective scale. To identify any problematic items that should be considered for deletion and ascertain the fit of each indicator to the appropriate latent variable, a one-factor CFA was conducted for each scale. By constraining each scale to one factor, the hypothesized model states that all items in that scale are describing the same latent variable. One item per latent variable was set to unity. This must be done to provide the latent variable with a metric equivalent to that of the indicators. Latent variables, by definition, are unobservable variables, and thus, have no scale. By matching the metric of the latent variable to the corresponding indicators, the scale of the latent variable becomes identified (Brown, 2006, pg. 62). Only complete data sets were included in the analyses; thus, list-wise deletion was used for any missing data. Separate CFAs were conducted using both pre- and post-semester data. All CFAs were performed using LISREL version 8.80 (Jöreskog & Sörbom, 2006).

In order to assess the degree to which the data fits the hypothesized model, many different fit indices must be used. There are two types of global fit indices commonly used in the literature to assess model fit: absolute and incremental. Absolute fit indices (i.e., chi-square, RMSEA, and SRMR) are estimates of how well an *a priori* model fits the data. Incremental, or comparative fit indices (i.e., TLI and CFI) reflect improvement of model fit compared to a

baseline model (Kline, 2011). The indices used in this study included Satorra-Bentler (SB) scaled chi-square (Satorra & Bentler, 1994), root mean squared error of approximation (RMSEA) (Steiger, 1990), non-normed fit index (TLI) (Tucker & Lewis, 1973), standardized root mean squared residual (SRMR), and comparative fit index (CFI) (Bentler, 1990). The SB scaled chi-square is a test for exact model fit, where the population covariances are fully reproduced by the hypothesized model. A non-significant result is desired and indicates that there are not significant discrepancies between the population covariances and those predicted by the model (Kline, 2011). However, the chi-square test is sensitive to sample size and will often produce a significant result for very small deviations in model fit. Thus, other descriptive test statistics are used to assess the fit of the model (Schermelleh-Engel, Moosbrugger, & Müller, 2003).

The RMSEA can range from 0 to infinity and is a measure of the approximate model fit in the population (Steiger, 1990). Because exact fit of the model in the population is impractical, the RMSEA is a measure of “close fit,” and in general, values < 0.05 are considered good and those < 0.08 are considered reasonable (Browne & Cudeck, 1992). The SRMR value ranges from 0 to 1 and is a “badness of fit” measure based on the standardized fitted residuals. By standardizing the residuals, the scale of the variables is taken into account (Schermelleh-Engel, Moosbrugger, & Müller, 2003). Hu and Bentler (1995) suggested that an SRMR value of < 0.05 is indicative of good fit and < 0.10 is acceptable. The TLI and CFI both take into account the chi-square values of the proposed model and the null baseline model (Brown, 2006). The TLI and CFI

values are normed and range from 0 to 1, with values ≥ 0.95 indicating good fit (Hu & Bentler, 1999). Only when several fit indices (both incremental and absolute) are considered together can the quality of model fit be assessed with reasonable propriety (Brown, 2006). Based on what is commonly accepted in the literature, the following cut-off values were used as an evaluation of acceptable model fit beyond the chi-square test statistic: RMSEA ≤ 0.05 , SRMR ≤ 0.10 , TLI and CFI ≥ 0.95 (Hooper, Coughlan, & Mullen, 2008; Hu & Bentler, 1999).

Component model fit was evaluated based on statistical significance ($p < 0.05$) and reasonable parameter estimates. The parameter estimates in this case refer to factor loadings and error variances. The model is specified by the researcher, but the parameter estimates are generated by the program used for analysis. Factor loadings refer to the degree that an indicator is linked to its corresponding latent variable. The error variance is specific to each indicator and reflects the specific variance associated with an indicator independent of all other indicators. Also included in the error variance is the measurement error for that indicator (Brown, 2006). In addition, modification indices were considered, when significant. The modification index (MI) is represented as a one degree of freedom chi-square statistic that estimates the difference between two nested models. Modification indices are parameter-specific and reflect the approximate decrease in the model chi-square statistic when the fixed parameter is allowed to be freely estimated (Brown, 2006). With regard to CFAs, MI values are most commonly associated with correlated residuals between two indicator variables or between an indicator and two factors in the model. Thus, the higher the MI, the

more likely it is that a particular indicator either is redundant or belongs with another factor. Modification indices are part of the evidence used to assess whether an indicator (item) should be dropped, aggregated with another indicator, or relocated to a different factor (latent trait). When the MI value exceeds the critical χ^2 value ($\alpha = 0.05$) using the degrees of freedom (*df*) from the model, then the appropriate modification should be considered (Hancock, 1999).

Multiple fit indices and other indicators of model fit are used collectively as a way of describing the overall quality of the model. Each fit index is different in what it describes, and no index is sufficient on its own to evaluate satisfactory model fit. When considered together, the decision to retain or re-specify a model can be more confidently justified.

Interview Participants

Interviews were conducted with students enrolled during the fall or spring semesters of a first-semester general chemistry course. Solicitation for interviews occurred in one of two ways: during administration of the scales, or by an announcement made in lab. Students completing the scales had the opportunity to indicate whether or not they would be willing to participate in an interview later in the semester. Initial contact, via e-mail, was made a few weeks after the scales were given, and all interested students were contacted. After the lab announcement, a sign-up sheet was made available for interested students to list their e-mail. These students were contacted shortly after the initial announcement was made.

Interview Design and Protocol

All interviews took place in a private interview room to ensure both participant confidentiality and audio quality. Prior to starting the interview, each participant was informed about the purpose of the study, the interview procedure, and the protocols for confidentiality. Following that, the participants were asked to sign a consent form approved by the IRB. By the time of the interviews, all participants had completed the scales and were provided a copy of their original answer choices for each scale. All interviews were audio recorded. A verbal probing interview approach was used; students were asked to read each item out-loud, explain the reasoning behind the answer choice they made, and comment on the readability of the items (Knafl et al., 2007). If a student's reasoning did not match their answer choice or was unclear to the researcher, additional probing questions were asked to clarify their interpretation of the item. This methodology is important in establishing the response process validity (Arjoon et al., 2013) for the modified items and response scales, ensuring proper readability and consistency between students' answer choices and reasoning among the target population (Barbera & VandenPlas, 2011).

Interview Data Collection and Analysis

All interviews were transcribed and coded for significant statements and emergent themes, each based on individual items (Creswell, 2013a). Items that showed consistent patterns of poor readability or multiple interpretations were flagged as candidates for removal from the scale. It is important that most participants agree on what an item means; otherwise, the significance of the

score for that item is diminished. Since validity is concerned with the inferences drawn from scores, the validity related to that item and its corresponding scale is threatened as well.

Quantitative-Qualitative Integration

Results from both qualitative and quantitative strands were integrated to support the most plausible set of items for each scale. The goal was to have evidence from the CFAs as well as the student interviews to support any changes in the scales. Depending on the item, this evidence may have been weighted differently. For example, if an item was detrimental to the model fit, but there was little qualitative evidence to support its removal, it was still considered for removal. In the same vein, if an item was unclear to many students or difficult to read, but the model did not improve significantly from its removal, this item was also considered for removal. In the best-case scenario, both qualitative and quantitative results showed that the item does not describe the latent trait adequately.

Phase two: Classroom-Based Study

Research Design

This phase of the study followed a quasi-experimental design whereby all data were quantitative. Three first-semester general chemistry sections were involved in this phase.

Quasi-experimental designs refer to studies that use a test variable (e.g., intervention or teaching strategy) across different groups, but do not include random selection of participants or random assignment of participants. As a

result, threats to internal validity must be carefully controlled or acknowledged, and the external validity, or generalizability of the results, is limited. Under the quasi-experimental umbrella are several additional categories of research designs. These include static-group comparison, nonequivalent control group, counterbalanced, and time-series designs (Martella et al., 2013). For the quasi-experimental portion of this phase, two different designs were used: a nonequivalent control group design and a static-group comparison. Two different designs were needed because several dependent variables were used in separate analyses.

The nonequivalent control group design was used to compare students in treatment and control conditions based on their scores for self-efficacy, effort beliefs, and interest. Nonequivalent control group design is used when the dependent variable is measured prior to and after introduction of the treatment. In this case, the scales were administered at the start and end of the semester with the intervention taking place in the middle of the semester. This design allows for initial test scores to be compared between the groups. As a result, any changes in students' scores can be more confidently linked to the treatment condition, and not to inherent differences between the groups from the beginning.

The static group comparison was used with the course performance data. Unlike the nonequivalent control group design, the dependent variable is only measured once for all groups in static group comparisons. Since course performance was only measured once at the end of the semester, this design is appropriate. However, the possibility of controlling for inherent group differences

in ability is lost. Thus, it is possible that students in one group might have had a higher chemistry ability level at the start of the semester than students in either of the other groups. This is a limitation that cannot be overcome without random assignment to groups. Random assignment was not possible in this context because students self-select into the course section of their choice.

Participants

Participants were recruited from three sections of first-semester general chemistry at a mid-sized university, located in the Rocky Mountain west, during the fall. Two of these sections were taught by the same instructor. As with phase one of the study, this sample was a convenience sample and cannot be considered representative of general chemistry students everywhere.

Scales and Demographic Form

All scales from phase one (self-efficacy, effort beliefs, and interest) were used during this phase of the study in their chemistry-specific adapted form. Further, any changes that were made to the scales based upon evidence gathered during phase one were retained. Thus, if any items were removed from the scales, they were not used for measurement during this phase of the study. Demographic information was also collected from all participants during this phase of the study. These items included: gender, age, declared major, time since high school chemistry, race/ethnicity, and whether or not they had taken college chemistry prior to this class.

Survey Data Collection

Survey data were collected at two time points (week 1 and week 13) during the semester from three sections of first-semester general chemistry. During the first week of the semester, an announcement was made by the researcher to all classes regarding the purpose of the study as well as relevant confidentiality information. Students were encouraged to participate, but were also made aware that participation was voluntary. Each student then received a packet with the survey, an answer sheet, and a consent form. Students indicated whether or not they consented to allowing their answers to be used by the researcher by answering “Yes” on the first item of the survey. In addition, students were asked to provide an identifier, so their sets of scores could be matched. Only students who consented and had initial scores were used in the analysis, thus consent forms were not given with the survey after the first data collection.

Performance Data Collection

Final percentage grades for all participants were obtained from the instructors of the course. In addition, American Chemical Society (ACS) Standardized General Chemistry Exam (first term 2010 version) scores were obtained. This exam was given as the final exam for the course.

Interventions

Three interventions were given throughout the semester at different time points, both task-based and lecture-based (see Appendix D for summary). These interventions only took place in the two sections taught by the same instructor,

and they were administered by the instructor. The instructor designated one section as treatment group A and the other as treatment group B. The control group was the third section taught by a different instructor, where nothing about the class was altered from normal. All in-class interventions in the treatment groups were equally desirable such that time on task and instruction remained constant between the two treatment groups.

The goal of these interventions was to test the efficacy of very brief discussions and activities on improving motivation and performance among college chemistry students. No published studies have been found that use these interventions in the context of college chemistry. Hence, this portion of the study was largely exploratory. The interventions are described below and occurred in chronological order.

The first intervention took place during the first week of classes, and was a written assignment aimed at encouraging students to consider what they value and why they value it. Values assessments have been correlated with positive performance gains among minority students and a reduction in the gender achievement gap in science (Cohen et al., 2006; Miyake et al., 2010). The values assessment chosen for this study was originally used by Cohen et al. (2006) in a study on seventh grade students, but was later used by Miyake et al. (2010) in a college physics course. The present study followed the protocol originally published by Cohen et al., 2006. Students in treatment group A selected two or three values most important to them from a list provided (e.g., being good at art, spiritual or religious values, music). Each student then wrote a short essay

reflecting on the most important value they chose. Students in treatment group B chose two or three values that were least important to them, and wrote a short essay on why the least important value they chose might be important to someone else. To reinforce the impact of the assessment, students in both groups were asked to give the top two reasons why the value they chose is either important to them (group A) or to someone else (group B). In addition, four statements were listed and students were asked to respond using a four-point Likert-type scale (strongly agree to strongly disagree). An example statement for treatment group A that was included is, "In general, I try to live up to these values"; and for treatment group B, "In general, some people try to live up to these values".

The second intervention occurred during the third week of classes. For treatment group A, it consisted of a 15-minute lecture by the instructor on the importance of having a growth mindset toward chemistry, including how the brain is malleable when learning. It has been shown that interventions concerning the malleable nature of the brain can significantly affect the trajectory of students' performance in school (Blackwell et al., 2007). For treatment group B, the instructor delivered a 15-minute lecture on study skills important for succeeding in chemistry. Following the lectures in both sections, students filled out a worksheet that directly linked with the lecture they were given. In treatment group A, students were asked to identify one activity they do well and then elaborate on how they learned and became better at that activity. In addition, students were asked to pair up with another student and describe their learning

process with their partner, then write down two things their partner did to become better at his or her activity. In treatment group B, all prompts were the same except the students were describing methods they use for studying chemistry instead of subjects or activities in which they do well. The purpose of the group A writing assignment was to encourage students to think about activities or subjects in which they do well, and likely value. The writing and discussion was intended to help students connect something they value with how they were able to increase their ability in that subject or activity. If students can make that connection, then they might be more likely to draw upon these experiences when confronted with the challenge of succeeding and doing well in chemistry. The writing activity in group B represented an equally desirable alternative to the group A writing activity that took a similar amount of class time.

The third and final intervention was an abridged version of the first intervention. This took place during week 7 of classes, prior to the second exam.

Pre/Post Data Analysis

All participants from the three sections in this phase of the study were included in the pre/post data analysis. Descriptive statistics and assumptions of normality were evaluated as described in phase one of this study. The time 1 and time 2 measurements were also evaluated in the same manner as described in phase one of this study with one addition. ANOVA tests were performed with race/ethnicity, gender, major, and treatment group as independent variables.

Multiple regression. Multiple regression analysis is a method used to predict or explain the variation in a dependent variable by examining its relations

to several independent variables (Pedhazur, 1997). By utilizing multiple regression analyses, the researcher is able to identify and separate the individual effects of distinct independent variables on a dependent variable. Multiple regression analysis was employed to examine to what extent self-efficacy, interest, and effort beliefs predict final grades. Multiple regression is the appropriate test for this application because the independent variables are continuous and the interest of the test is prediction. Separate regression analyses were run for each treatment group as well as the control group. As with the other analyses in this study, each participant's item scores were aggregated to produce a mean score for each scale. Prior to the multiple regression tests, bivariate correlations between all scale scores, final grade, and ACS score were obtained. Bivariate correlations represent the degree to which two variables vary together, or co-vary (Martella et al., 2013). For example, if people wear shorts more when it is warm outside, the outside temperature and frequency of shorts wearing would be said to co-vary. Since there is no precedent to this research in which all three motivational variables have been measured in a single study, the bivariate correlations were used as a guide for which variables to include in the regression model. Variables that correlate significantly with final grade were added as predictors to the preliminary regression model. After running the initial models, the results were checked for multicollinearity.

Multicollinearity is a term that is used to describe strong correlations among the independent variables in the model. If two or more variables are explaining much of the same variance, due to their inherent redundancy, the

results from the multiple regression analysis cannot be trusted (Pedhazur, 1997). Often, multicollinearity results in an inflation of Type I error, or failing to reject the null hypothesis when in fact, it should be rejected. This means that predictors that are significant in the model are shown not to be by the test results. To prevent multicollinearity, variables should be distinct enough to not overlap much in their explained variance of the dependent variable. Several diagnostics are used to check for multicollinearity in the model. The one that was used in this study is the variance inflation factor (VIF). Variance inflation factor indicates the increase in the variance of a predictor coefficient, b , when it is highly correlated with another predictor in the model. VIF values can range from 1 to infinity, with larger VIF values indicating a detriment to the model. No cutoff value for VIFs has been agreed upon by researchers (Belsey, 1991), thus the VIFs should be evaluated as relative to one another (Pedhazur, 1997). Hence, a relatively large VIF compared to other variables in the model may indicate a problem with that variable.

Assumptions for the models were verified including multivariate normality, linearity, homogeneity of variance, and normal distribution of the residuals. In addition, tests for outliers were used. Multivariate normality was considered by looking at the skew and kurtosis values of each aggregated scale. Skew and kurtosis values outside of a range of -1 to +1 were examined further. Linearity and normal distribution of the residuals were examined by plotting the standardized residuals (y-axis) versus the standardized predicted values (x-axis). If the points are randomly distributed, these assumptions can be upheld (Pallant,

2010). Similarly, the homogeneity of variance (or homoscedasticity) assumption can be made by looking at a plot of the standardized residuals (y-axis) versus the standardized predicted values (x-axis). If the points are scattered randomly, this assumption is upheld, meaning that the variance of responses are relatively uniform throughout the data set (Pallant, 2010).

There are several diagnostics used to test for potential outliers and influential points that were considered: leverage, Cook's D (distance), and standardized DFBETAs (DFBETASs). Leverage is a value that indicates the influence of a case on the independent variables only and cannot exceed 1. High leverage values are indicative of an influential case. A leverage value is considered high when it exceeds $2(k + 1)/N$, where k is the number of independent variables in the model and N is the sample size (Pedhazur, 1997). Cook's D is a measure that can indicate if a case is influential on the independent variable(s), dependent variable, or both. As with leverage, high Cook's D values are not desired, and they can range from 0 to infinity. Significance tests for Cook's D exist, but it is more realistic to consider Cook's D as a relative value in the data set. In other words, the significance tests may indicate a case is not an outlier, but in fact, it could be very influential on the model (Pedhazur, 1997). Both Cook's D and leverage are global indices, by detecting a possible influential case, but they do not describe what the effect of removing the influential case would be. However, standardized DFBETAs fill in the gap by estimating the change in the standardized regression coefficients when a potential influential case is removed. A large change in the regression coefficients upon removal of a

case indicates that the case is rather influential on the model. A general cutoff for DFBETASs has been proposed by Mason, Gunst, & Hess (1980) as $3/\sqrt{n}$, where n is the sample size. Cases that generate values higher than this cutoff should be considered as potentially influential.

This set of assumptions was checked for the preliminary models and for all subsequent models that were run. Any severe violation of these assumptions that was detected rendered the results of that particular model untenable and it was subsequently discarded.

Path analysis. Following multiple regression analysis, a set of *a priori* path models was tested using the data gathered on students' self-efficacy, interest, effort beliefs, and final grades. The difference between the regression models and path models is the interpretation of the effects of independent variables and causality. The regression coefficients of independent variables can be interpreted as affecting the value of a dependent variable. What cannot be described in regression analysis are the effects of one independent variable on another in addition to the effects on the dependent variable. Path analysis allows the researcher to generate a model that will estimate the effects of independent variables on each other as well as the dependent variable and the causal directions associated with these effects (Pedhazur, 1997).

No studies have been found which link self-efficacy, effort beliefs, and interest together with academic performance. However, several studies have been conducted, as pointed out in the literature review, which examine these constructs individually with performance, and in some cases with one of the other

constructs studied herein. Based on these studies, several models were proposed and tested. However, due to a dearth of research and theory that connects all of the motivational constructs to be examined herein, this path analysis was exploratory in nature.

Intervention Data Analysis

The main focus of this analysis was to investigate differences between the treatment and control groups. Three interventions were given and self-efficacy, interest, and effort beliefs were measured twice in all three sections. Only students who provide data from all five data collections (for the treatment groups), and both data collections for the control group as well as final course percentages were included in this portion of the analysis. Although the interventions are targeting effort beliefs and course performance, differences in self-efficacy and interest were examined as well.

MANCOVA. One-way multivariate analysis of covariance (MANCOVA) was used to test whether students differ on any of the aforementioned variables, based on the group to which they belong. The groups in this case refer to the different course sections included in this phase of the study. MANCOVA is a statistical method used when comparing groups on more than one dependent variable, whereby concomitant independent variables can be controlled. This is a suitable test here because performance (course grade) and all three motivational variables at time 2 are dependent variables. The time 1 measured variables were treated as covariates to account for individual differences that students had before the course began. In this way, any group differences found at the end of

the semester can be more reliably attributed to events and changes that took place during the semester. The three groups used in this phase of the study (treatment A, treatment B, and control) were tested for differences on all three motivational variables and course grade. MANCOVA differs from ANCOVA by the fact that ANCOVA is appropriate only when there is one dependent variable. One MANCOVA is preferable to performing several ANCOVAs because Type I and Type II error rates can be inflated with multiple statistical tests using the same sample (Haase & Ellis, 1987; Pallant, 2010). MANCOVA can be thought of as a composite of several ANCOVAs, and *post-hoc* tests for MANCOVA can reveal which dependent variable(s) the subjects differ. Hence, a significant MANCOVA result can be used as a springboard for testing the differences among groups on individual dependent variables by performing ANCOVAs without inflating Type I error (Haase & Ellis, 1987). MANCOVA is a way to focus the statistical tests to those that are shown to be significant instead of testing indiscriminately.

Several assumptions must be met for MANCOVA tests to be tenable, and all were tested in this study prior to the interpretation of any results. These include multivariate normality, tests for outliers, linearity, multicollinearity, and homogeneity of covariance (Pallant, 2010). Multivariate normality was tested by using the Mahalanobis distance. The Mahalanobis distance is the distance of a particular case from the centroid of all other cases, where the centroid is generated from the means of all variables (Tabachnick & Fidell, 2007). It indicates whether a case shows a unique pattern of scores across all of the

independent variables when compared to the rest of the data set. Once the Mahalanobis distance has been calculated for each case, each is compared to a critical χ^2 value based on the number of dependent variables, and with an alpha value of 0.001 (Pallant, 2010). Values greater than the critical χ^2 value were examined further as potential outliers and considered for removal from the data set. Linearity, in the context of MANCOVA, means that each pair of dependent variables generates a straight line when plotted against each other. Thus, plots of each pair of dependent variables were generated and examined for a linear trend. Multicollinearity, as mentioned above, indicates high correlation between two variables in a model. In the case of multiple regression, the variables of concern are independent variables, but with MANCOVA the variables of concern are dependent variables. To check for multicollinearity in MANCOVA, a table of bivariate correlations was generated and any values greater than 0.8 were considered highly correlated. Variable(s) with high degrees of multicollinearity were examined in terms of the degree of detriment to the model. Those that were deemed detrimental to the model were considered for removal from the analysis. The homogeneity of covariance assumption was tested using Box's Test of Equality of Covariance Matrices (Box's M). This test is used to indicate whether the covariance for the variables in the model is homogeneous, or similar, across all groups. A non-significant result is desired, and the alpha value used for the test is 0.001 (Pallant, 2010).

The test for outliers was conducted using the outlier labeling method (Hoaglin & Iglewicz, 1987; Hoaglin, Iglewicz, & Tukey, 1986; Tukey, 1977). This

method uses the interquartile range to generate useful maximum and minimum cutoffs of values for the variable in question. The interquartile range is the middle 50% of the data points in a normal distribution. It ranges from the 25% mark of the data points to the 75% mark, the difference of which equals 50%. To calculate the cutoff values, the interquartile range is multiplied by 2.2 (Hoaglin & Iglewicz, 1987). To generate the minimum cutoff, this value was subtracted from the minimum value in the interquartile range (25%); and to generate the maximum cutoff, this value was added to the maximum in the interquartile range (75%). The result was an extension of the range of values for a given variable, and anything outside of that range was considered an outlier.

Following the acceptance of all assumptions for MANCOVA, the MANCOVA was tested for significance. There are four major test statistics used by researchers to determine whether the MANCOVA model is significant: Wilks' Lambda, Pillai's Trace, and Hotelling's Trace, and Roy's Largest Root (Pallant, 2010). Wilks' Lambda has been recommended for general use, and was used in this study (Tabachnick & Fidell, 2007). If the MANCOVA model is significant, then the individual ANCOVAs will be analyzed for each of the dependent variables.

Prior to the ANCOVA analysis, the assumption of homogeneity of variance must be met. To test for the homogeneity of variance assumption, Levene's test was employed. As with Box's M test, a non-significant result is desired and indicates that the variance is similar for all groups with respect to a particular dependent variable. Following a non-significant result for Levene's test, the

individual ANCOVAs were examined for significant differences among the dependent variables between the two sections of students. *Post-hoc* tests were performed for any significant ANCOVAs found. As with the time 1 and time 2 measurements described above, Tukey's tests were used for *post-hoc* analyses.

Phase Three: Effort Beliefs Interviews

To investigate the sources, influences, and reasons for changes in students' effort beliefs in chemistry, a series of student interviews was conducted by the researcher.

Participants

The interview participants were students who took first-semester general chemistry in either of two consecutive fall semesters. The effort beliefs scale was given at the start and end of the semester both years. Some of the students were interviewed in the fall semester and others were interviewed in the following spring. Students were recruited via email during the first few weeks of the semester and were informed that participation was voluntary. For those who were recruited for interviews during the fall, two groups of students were of particular interest to the researcher and were purposely sampled: those with high effort beliefs average scores (> 4.5) and those with low average scores (< 3.0). For participants in the spring semesters, three groups of students were of primary interest to the researcher and were purposely sampled: those with a relatively large drop in their effort beliefs scores across the semester (> 0.8), those with a relatively large increase in their mean scores (> 0.8), and those with high initial effort beliefs scores with no change in their scores despite a low final

grade in the course. Students who report higher or lower effort beliefs scores at the end of the semester, compared to the beginning, may provide valuable insight as to what caused their scores to change. In addition, students who receive low grades in the course, but report no changes in their positive effort beliefs are interesting cases because it is expected that performance and effort beliefs would be positively correlated. Thus, interviewing these students helped to broaden the understanding of effort beliefs and its permanence among students.

Interview Design and Protocol

The purpose of these interviews was to explore the sources and influences of college-aged students' effort beliefs toward first-semester general chemistry. In addition, the reasons for changes in students' effort beliefs across a semester of general chemistry was investigated. Students who participated during the spring had already completed the effort beliefs scale at the beginning and end of the previous semester. Those who participated in the fall semester had, at the time, only completed the effort beliefs scale at the beginning of the semester. In both cases, the researcher had a copy of the students' scores ready for a discussion prior to the interview.

The interviews took place in a private interview room to minimize disturbances and secure confidentiality. The interviews began with the researcher addressing the purpose and confidentiality of the interview, and presenting a consent form for the participant to sign. Following that, a series of questions were asked regarding the participant's academic background including

their experiences in school, experiences with chemistry, and where their effort beliefs came from. For the fall interviews, the researcher then focused on what the participants were thinking of when they filled out the survey and what caused them to make the answer choices they did. For the spring interviews, the subsequent questions mainly focused on their experience in first-semester general chemistry, including how they viewed their effort and their performance in the class. They were then asked for their reasoning behind any changes in their scores or why their scores stayed the same. A semi-structured approach was used throughout the interview. Although most of the questions were structured prior to the interview, the researcher encouraged the participants to elaborate beyond the set of questions during certain times of the interview. By structuring only some of the questions, the interviews were partially standardized between each participant, but not so limiting that the participants were unable to guide the discussion at times. In addition, by not structuring all of the questions, the researcher reduces the risk of asking leading questions (Shank, 2002).

Interview Data Collection and Analysis

All interviews were audio-recorded and transcribed by the researcher. The audio files and transcripts were given a four-digit numeric code and stored on a password-protected computer to ensure confidentiality. Only the researcher had access to the list corresponding the numeric code with the name of the participant. The interview transcripts were coded for significant statements and emergent themes based on the students' experiences and performances in general chemistry, as well as the sources and influences of their effort beliefs

prior to and during the chemistry course (Creswell, 2003). Common themes and phrases among the participants were summarized and then checked against the transcripts again to ensure fidelity and clarity.

CHAPTER IV

ARTICLE 1: ANALYSIS OF STUDENTS' SELF-EFFICACY,
INTEREST, AND EFFORT BELIEFS
IN GENERAL CHEMISTRY

Reprinted (adapted) with permission from Brent Ferrell and Jack Barbera, Analysis of students' self-efficacy, interest, and effort beliefs in general chemistry. *Chemistry Education Research and Practice*, 2015, 16, 318-387. doi: 10.1039/C4RP00152D Copyright (2015) Royal Society of Chemistry. This manuscript cannot be reprinted or copied without permission from the publisher.

Abstract

Research in academic motivation has highlighted a number of salient constructs that are predictive of positive learning strategies and academic success. Most of this research has centered on college-level social sciences or secondary school student populations. The main purpose of this study was to adapt existing measures of personal interest and effort beliefs to a college chemistry context. In addition, a chemistry-specific measure of self-efficacy was evaluated in a modified form. This set of scales was initially administered at two time points in a first-semester general chemistry course to a sample of undergraduates ($n_1 = 373$, $n_2 = 294$). Confirmatory factor analyses (CFA) were conducted to determine whether the scales were functional in a chemistry context. Following revision of the scales, all CFA models demonstrated acceptable fit to the data. Cross-validation of the revised scales was performed using two different populations

($n = 432$, $n = 728$), with both studies producing similar model fits. Furthermore, our data shows that chemistry majors reported higher self-efficacy and interest than non-science majors. Cronbach's alpha estimates ranged from 0.75 to 0.92 for the revised scales across all studies. This set of scales could provide useful tools for assessing general chemistry students' motivation and the motivational impacts of various teaching practices.

Introduction

Introductory chemistry is a course required by several disciplines. At most institutions, enrollment overwhelmingly consists of students outside of the discipline of chemistry. Often, many students struggle through chemistry and are unsuccessful due to the complexity and abstract nature of the content (Nakhleh, 1992). The combination of content difficulty and the fact that most students are fulfilling a credit requirement for their non-chemistry majors generates an interesting classroom environment for the introductory-level chemistry course. Many factors can influence whether a student is successful in chemistry. There are some obvious characteristics of students including: inherent aptitude and prior experience in chemistry, which can be predictive of success in chemistry (Tai *et al.*, 2005). However, research has shown that cognitive factors such as these are not sufficient to predict achievement, but must be augmented by adaptive motivational processes (Dweck, 1986; McCoach & Siegle, 2003; Zusho *et al.*, 2003).

The importance of motivation for learning and achievement in any classroom setting is indisputable (Dweck, 1986; Schunk, 1991; Ames, 1992; Hidi & Harackiewicz, 2000; Singh *et al.*, 2002; Zusho *et al.*, 2003). Student motivation has been described as a theoretical construct that can explain “the degree to which students invest attention and effort in various pursuits” (Brophy, 2010). Motivation and ability are the two major components for academic success among students (Hidi & Harackiewicz, 2000). Despite this, students’ motivation can easily be ignored due to its complexity, or oversimplified as an unchanging

facet of one's character. Motivation in the classroom is very complex, and can fluctuate in different situations (Nicholls *et al.*, 1989; Pintrich, 2003) and among different subjects (Guay *et al.*, 2008). Students' motivation toward school tends to become highly differentiated throughout grade school, as individuals encounter various situations and experiences that shape their interests and conceptions of ability. As such, it is important for researchers to study motivation in specific contexts (Pintrich, 2003).

Although much research has been conducted to understand the motivational and affective factors that influence performance and student engagement in the college classroom, only a small fraction of this has been directed towards science classrooms. In 2012, the National Research Council called for a collective effort on the part of science education experts across many disciplines to put together a report that would highlight current research areas in education, as well as areas that are lacking across these disciplines. As pointed out in this Discipline-Based Education Research (DBER) report, students' dispositions and motivations to learn science and engineering are largely understudied and are of "central importance" (National Research Council, 2012). Nevertheless, many researchers have explored specific motivational constructs and processes in the context of college chemistry both prior to and since this report (Dalgety *et al.*, 2003; Zusho *et al.*, 2003; Bauer, 2005; Dalgety & Coll, 2006; Taasoobshirazi & Glynn, 2009; Uzuntiryaki & Aydin, 2009; Xu *et al.*, 2013; Villafane *et al.*, 2014).

Our aim in this study is to expand upon the current body of research directed at understanding motivational processes among students in college chemistry. In particular, we were interested in modifying existing measures of motivational processes for use in college chemistry. We chose to examine three distinct constructs that have been linked to motivation in students: self-efficacy, interest, and effort beliefs (Zimmerman, 2000; Hidi & Renninger, 2006; Jones *et al.*, 2012). We chose these three variables both for their salience within motivation literature and their influence on student performance and retention. Self-efficacy has been found to be positively correlated to achievement outcomes in many studies, as well as the adaptive motivational processes, effort and persistence (Pintrich & De Groot, 1990; Multon *et al.*, 1991; Pajares & Miller, 1994; Zusho *et al.*, 2003). Individual interest is less consistently correlated with performance, but is more strongly linked to positive learning strategies, such as mastery goals and attention (Hidi & Renninger, 2006). Furthermore, discipline-specific individual interest has been positively correlated with choice of major and number of classes taken within that discipline (Harackiewicz *et al.*, 2008). Positive effort beliefs are very highly correlated with an incremental theory of intelligence, which states that competence is not fixed, but malleable (Blackwell *et al.*, 2007; Dweck, 2012; Jones *et al.*, 2012). Effort is absolutely vital for success in any college classroom. However, preceding the action of effort itself must be a positive, adaptive belief about the potential impacts of the action. Hence, effort beliefs are indicators about how a student perceives the impact of their effort on learning and performance, and students who believe their

competence can be changed through effort tend to be more motivated and perform at higher levels than those who do not (Dweck, 2000).

The emphasis in academic motivational research has shifted from behavioral aspects such as drive and reinforcement to beliefs, goals, and expectations over the last 40 years (Wigfield & Eccles, 2002b). These modern theories of motivation point to the critical role that expectancies and beliefs play in adaptive learning patterns among students (Eccles and Wigfield, 2002; Zeldin *et al.*, 2008). By researching these beliefs among students in college chemistry, we can have a more clear understanding of the basis for motivational processes that exist in our classrooms. However, to do this we first need measures of the various motivational aspects that have been tested within the target population.

Self-Efficacy

Self-efficacy is rooted in social cognitive theory and is defined as the self-appraisal of one's capacity to execute a specific task (Bandura, 1977). Efficacy beliefs or expectations are self-referent and guide an individual toward certain actions and away from others (Pajares, 1996). Self-efficacy must be distinguished from two theoretically similar constructs, outcome expectations and self-concept. Outcome expectations are beliefs that certain behaviors will lead to certain outcomes. Both self-efficacy and outcome expectations influence motivation, but self-efficacy is thought to play a larger role in predicting achievement (Zimmerman, 2000). While both self-efficacy and outcome expectations are task-specific, self-concept is much more broadly defined, and has more to do with one's beliefs about their self-worth and competence within a

domain (Pajares & Miller, 1994). Moreover, self-concept has been defined in many different ways, is tied to affective as well as cognitive judgments, and is inherently norm-referenced (Hansford & Hattie, 1982; Bong & Clark, 1999). In contrast, the prevailing definition of self-efficacy has remained virtually unchanged since its inception (Bandura, 1977). Also, self-efficacy judgments are less likely to be influenced by social comparisons and affective swings due to the task-specific, objective nature of the construct (Bong & Clark, 1999). For these reasons, and the notable presence of self-efficacy in the literature, we chose to measure self-efficacy instead of other related constructs.

Bandura (1986) argued that efficacy expectations are drawn from four sources of information: performance accomplishments, vicarious experiences, verbal persuasion, and physiological states. The most dominant of these sources are performance accomplishments, because they are founded upon personal mastery experiences (Bandura, 1977). However, some suggest the relative salience of sources for self-efficacy beliefs may be different for males and females (Zeldin *et al.*, 2008). Nevertheless, a student in chemistry is more likely to have high efficacy expectations for a particular task if he or she has already successfully completed that task. When asked to explain their self-efficacy in college chemistry, students noted their prior success in chemistry as a common theme (Dalgety & Coll, 2006a). From a quantitative approach, Lopez and Lent (1992) found that prior experience in math explained the most variance in self-efficacy scores when considering math self-concept, interest in math, and

perceived value of mathematics. Equally important to considering the sources of self-efficacy, is the influence that self-efficacy has on the student.

Self-efficacy is hypothesized to have far-reaching implications in academics by influencing students' effort, perseverance, and emotional reactions to specific tasks in school (Lent *et al.*, 1984; Lopez & Lent, 1992; Pajares & Kranzler, 1995; Pajares, 1996). Lent *et al.* (1984) investigated the relationship of students' self-efficacy beliefs to persistence in technical and science majors. They found that students who reported higher self-efficacy scores for completing their educational requirements were more likely to persist in their major. This supports Hackett and Betz's (1981) hypothesis that self-efficacy is linked to persistence in career goals.

Other studies across several disciplines suggest that self-efficacy is related to academic performance, problem solving, college entrance, and college major choice (Betz & Hackett, 1983; Lent & Hackett, 1987; Lopez & Lent, 1992; Pajares & Kranzler, 1995; Andrew, 1998; Britner & Pajares, 2001; Schunk & Pajares, 2002; Zusho *et al.*, 2003; Parker *et al.*, 2014). Zimmerman, Bandura, and Martinez-Pons (1992) found that students' academic self-efficacy scores significantly predicted ($\beta = .39$) their final grade in a high school social studies class. In a college setting, it was found that students' self-efficacy beliefs were significantly correlated to their grade-point average (GPA) and accounted for more variation in GPA than ACT scores (Gore, 2006).

Generalization can also occur with efficacy expectations, such that a student who has experienced mastery with one task could report high efficacy

expectations for a similar task (Bandura *et al.*, 1969). However, the degree of generalization is limited to the domain of functioning (Zimmerman, 2000). For example, one cannot assume that just because a student has high self-efficacy in biology, he or she will also have high self-efficacy in chemistry. Thus, it is important, when measuring and describing students' self-efficacy, to ensure that it is domain-specific.

Chemistry has been referred to as the “central science” and it is believed that mastery of chemistry concepts is influential for success in later science courses (Tai *et al.*, 2005). Although chemistry course requirements are not as widespread as those of math, there are many professions unrelated to chemistry, particularly in health care, that require a background in chemistry (Brown *et al.*, 2012). In fact, students pursuing degrees other than chemistry make up the bulk of the enrollment in introductory chemistry at larger universities, and it is these students who are least likely to exhibit high self-efficacy in chemistry (Uzuntiryaki & Aydin, 2009). Therefore, since self-efficacy could have a substantial influence on students' achievement and retention in college chemistry, it is an important factor to consider in chemical education research (Zusho *et al.*, 2003).

Several researchers have investigated chemistry-specific self-efficacy in university-level classes (Dalgety *et al.*, 2003; Zusho *et al.*, 2003; Taasobshirazi & Glynn, 2009; Uzuntiryaki & Aydin, 2009; Villafane *et al.*, 2014). Taasobshirazi and Glynn (2009) conducted a study where undergraduate introductory chemistry students were asked to solve a series of problems along with judging their self-efficacy in chemistry. It was found that students who reported high self-efficacy

were more likely to use forward-working strategies and obtain the problem solution than students who reported low self-efficacy. In a related study, Zusho et al. (2003) investigated the role of several motivational processes on achievement in chemistry, as well as the correlations between them. They found that self-efficacy was the best predictor of final course grade, when measuring SAT-math, task value, and rehearsal strategies. Both of these studies support Bandura's (1986) notion that self-efficacy is positively related to achievement through persistence and effort. Villafañe et al. (2014) explored student self-efficacy trajectories across a semester of preparatory chemistry. Their results suggest that individual characteristics (race/ethnicity and gender) could influence the degree to which students show an increase in self-efficacy across the semester. Although these studies have addressed important gaps in the literature, more research is needed that investigates the interplay between different motivational variables.

Interest

Interest, in academic settings, can be described as a psychological state where the student is engaging or has a predisposition to engage with content over time (Hidi and Renninger, 2006). Hidi and Baird (1986) argued that interest is more than "arousal," but must be considered as a process. As a process, interest is said to endure and persist through time. There are two types of interest outlined in the literature: situational interest and individual interest (Hidi, 1990). Situational interest refers to an interest that is triggered spontaneously through interaction with the environment (Harackiewicz *et al.*, 2008). This type of interest

can be considered to deliver a sense of enjoyment and curiosity, but may or may not persist within the person (Renninger, 2000). Individual interest refers to a relatively stable interest that has developed over time and is associated with an enduring predisposition for the person to reengage with specific topics, subject areas, or activities (Hidi, 1990; Schiefele, 1991; Hidi & Renninger, 2006).

Individual interest can be further broken down into two components: feeling-related and value-related (Schiefele, 1999). Feeling-related interest (e.g., Chemistry is fascinating to me) is tied to stimulation and enjoyment. Value-related interest (e.g., The material we are learning in chemistry is important for me to know) is associated with importance and personal significance and has been positively correlated with performance in academic contexts (Hulleman *et al.*, 2008).

Hidi and Renninger (2006) postulated that interest is a developmental process, which occurs in four phases: triggered situational interest, maintained situational interest, emerging individual interest, and well-developed individual interest. Triggered situational interest results from temporary cognitive or affective changes in the individual (Hidi & Baird, 1986). Maintained situational interest occurs after a triggered event, where the individual's interest is "held" and involves focused attention that endures for a period of time. If maintained situational interest persists long enough, it becomes emerging individual interest, characterized by stored value, positive feelings, and consistent reengagement with the material or activity. At some point, emerging interest can become well-developed interest following substantial reengagement with the material.

Individuals with well-developed interest seek out answers to questions, and are likely to be resourceful when answers are not easily found. Also, it is possible for these individuals to expend effort, but feel as though the task is “effortless” (Renninger & Hidi, 2002). Although the development of interest is not the same for all individuals, Hidi and Renninger (2006) argued that there is no evidence suggesting well-developed interest can spawn without the individuals first being exposed to the area and experiencing triggered interest.

In the same vein as efficacy beliefs, interest is content-specific and represents a personal significance between the individual and the object of his or her interest (Schiefele, 1991; Renninger & Hidi, 2002). Science, as a subject matter in schools, is somewhat broad and diffuse in elementary and middle school years, but becomes differentiated and more focused during high school and college years. In light of this, several studies on interest in science and/or students’ perceptions of the value of science at lower grade levels have focused on “science” as a whole (Anderman & Young, 1994; Singh *et al.*, 2002; Tuan *et al.*, 2005). However, studies on interest in science dealing with high school and college student populations tend to center on specific disciplines, such as chemistry, physics, and biology (Dalgety *et al.*, 2003; Uitto *et al.*, 2006; Gungor *et al.*, 2007; Barbera *et al.*, 2008). These studies underscore the importance of developing measures to target individual interest as a domain-specific construct. Hence, to understand more about interest as a component of motivation in the college chemistry classroom, the focus of the interest measure must be specifically directed towards the discipline of chemistry.

Effort Beliefs

Effort can be considered as part of the attributional theory of motivation, and from this theoretical standpoint, is intimately tied to conceptions of ability (Weiner, 1985). Weiner (1985) pointed out that ability and effort are the most salient causal ascriptions to achievement. In short, students believe that those who have high ability and display high effort will be more successful than students who have low ability and display low effort. The notion that effort and persistence has a positive effect on the academic outcome of a student has been supported by empirical studies (Stipek & Gralinski, 1996; Elliot, 1999). For example, Elliot (1999) found that self-reported persistence and effort were positive predictors of academic performance. Effort was found to be a mediator between adaptive mastery goals and academic performance. Another study revealed that students who expressed positive beliefs toward the value of effort do not necessarily show increased performance, but do tend to focus more on mastery and the development of their abilities (Stipek & Gralinski, 1996). Effort is certainly a key component in the academic success of students. It must be considered when making judgments about academic performance, due to how it mediates the link between motivational constructs and academic outcomes (Elliot, 1999; Goodman *et al.*, 2011)

Students' beliefs about effort are a precedent to effortful actions and are highly correlated with their conceptions about intelligence and ability (Blackwell *et al.*, 2007; Jones *et al.*, 2012). These conceptions are referred to as implicit theories of intelligence. Implicit theories are "beliefs about the nature of human

attributes”; two implicit theories of intelligence have been described in the literature: incremental and entity (Dweck, 2012). An incremental theory of intelligence is characterized by the view that intelligence is malleable and can change over time with the expenditure of effort. Conversely, individuals who hold an entity view of intelligence see it as fixed and unchanging, independent of effort (Dweck & Leggett, 1988). Students holding an incremental view of intelligence are more likely to see effort as enhancing ability and apply more effort to overcome obstacles. On the contrary, those endorsing an entity view of intelligence are less likely to put forth effort in the face of failure, less likely to be interested in a subject, and exhibit achievement gaps when compared to incremental theorists (Dweck & Sorich, 1999; Hong *et al.*, 2004; Dweck, 2012). Effort beliefs and implicit theories of intelligence are fundamentally different constructs, but deeply related (Blackwell *et al.*, 2007; Jones, Wilkins, Long, & Wang, 2012). Between these two, the vast majority of research has centered on implicit theories, leaving a gap with respect to effort beliefs. No studies were found on effort beliefs in a college setting, but a few studies have investigated effort beliefs in secondary school (Blackwell *et al.*, 2007; Jones *et al.*, 2012).

It is important that students feel they can improve upon their abilities with persistence and effort. This could be a challenge to get across to college students, as it is likely that their beliefs about intelligence and effort have already developed by the time they reach the college classroom. However, if we understand more about effort beliefs and how they can change among college-age students, new instructional strategies can be implemented so that more

students will endorse positive effort beliefs in their classes. This is particularly important in math and science, because most students have experienced both prior to coming to college, and have likely developed beliefs about their abilities in those domains.

Purpose and Rationale for Current Study

There is a need in the chemistry education community to understand some of the motivational and affective components of students enrolled in chemistry courses. What is true of one college subject is not necessarily true of another, and given the lack of motivational research in college-level sciences, it is important that this need be addressed.

In order to effectively assess a large classroom of students on their motivational characteristics and dispositions, instructors and researchers must rely upon easy to administer self-report scales or instruments, consisting of items targeted at measuring a specific latent trait or group of traits. As with any scale or instrument, either in the physical or social sciences, the quality of data that can be produced from it depends largely on the quality of the data generated with the target population (Barbera & VandenPlas, 2011; Heredia & Lewis, 2012). Thus, steps must be taken to ensure that a scale or instrument will produce valid and reliable results when used with the target population.

While general and science-specific motivation instruments exist, such as the Motivated Strategies and Learning Questionnaire (MSLQ), and the Science Motivation Questionnaire (SMQ), the current availability of individual scales that measure motivational variables in college chemistry is limited (Pintrich *et al.*,

1993; Glynn *et al.*, 2009). In 2003, Bauer and colleagues published the Chemistry Self-Concept Inventory (CSCI), a 40-item survey that was adapted to measure students' self-concept in five domains: math, chemistry, academic, academic enjoyment, and creativity. There are two instruments (Colorado Learning Attitudes about Science Survey (CLASS) and Chemistry Attitude and Experiences Questionnaire (CAEQ)), which measure interest in chemistry (Dalgety *et al.*, 2003; Barbera *et al.*, 2008). The MSLQ and SMQ also have general academic interest scales, which could be adapted for a chemistry context. However, none of these instruments were designed based on the prevailing theoretical underpinnings of interest theory from educational psychology (Schiefele, 1991; Renninger, 2000). The CAEQ, being the largest instrument for motivation and affect in college chemistry, also has a self-efficacy scale. The CAEQ together with the College Chemistry Self-Efficacy Scale (CCSS) represent the only two instruments designed specifically to measure self-efficacy in a college chemistry setting. Both of these were designed using Bandura's (1986) theory of self-efficacy, which is widely accepted among researchers across many disciplines. The MSLQ and SMQ also have self-efficacy scales. However, the items in both scales are very general, relating more to the course as a whole than to tasks within the course. As mentioned above, self-efficacy is conceptualized as confidence at the task-level. Thus, scales purported to measure self-efficacy in a particular academic domain should be written with items targeting specific tasks encountered within that academic domain. We carefully examined the self-efficacy items from those that had been

used in chemistry and related disciplines for wording that was most appropriate for an introductory chemistry class in a college setting. The CCSS contained items that were relevant, task-specific, and readable for college students in introductory chemistry. No chemistry-specific measures for effort beliefs, or implicit theories of intelligence were found.

The goal of the present study was to identify, utilize, and evaluate existing measures of motivational constructs (initial interest, maintained interest, self-efficacy, and effort beliefs) in a college chemistry setting. As doing this required modification of items and scales, this manuscript represents evidence of validity and reliability for these measures prior to their use in future studies of instructional styles. Items from existing scales must be modified to be discipline-specific, as the constructs under investigation are operationalized around a meaningful connection between the subject area and the individual (Bandura, 1986; Schiefele, 1991). Modifying items, scales, and instruments for subject-specific language and meaning is common practice in science and math education (Barbera *et al.*, 2008; Taasobshirazi & Glynn, 2009; Jones *et al.*, 2012). However, any modifications to an item, scale, or instrument must be followed up by an investigation for validity and reliability evidence (Barbera & VandenPlas, 2011).

Items were taken from three independently published scales, each measuring a separate motivational construct (self-efficacy, interest, or effort beliefs) in academic settings. The items from the effort beliefs and interest scales were adapted to fit a college chemistry context. The self-efficacy items were

originally developed for college chemistry and did not need modification. All scales were subjected to confirmatory factor analysis to determine if the structure of each scale matched that proposed by the individual scale developers. In addition, indicators of global and component model fit were used to assess whether any of the scales should be modified. Student interviews were also conducted and used as validity evidence for potential modifications of items and scales. Interviews with the target population are a vital part of evaluating an item, scale, or instrument for response process validity (Arjoon *et al.*, 2013; Wren & Barbera, 2013). The data collection and subsequent analyses were guided by the following research questions:

- (1) What modifications are needed to produce brief, chemistry-specific scales of self-efficacy, interest, and effort beliefs?
- (2) What evidence supports the functioning of each of the modified scales?
- (3) To what extent do students' interest, effort beliefs, and self-efficacy change across a semester of college chemistry?

The current study will contribute to the growing body of literature centered on motivational and affective processes among students in college chemistry courses. In addition, brief chemistry-specific scales for measuring three salient motivational beliefs will be made available to educators and researchers interested in gauging the motivational climate of their chemistry classrooms. These scales will be used in a follow-up study as variables in a path analysis to investigate a set of *a priori* motivation models that will include measures of academic performance in general chemistry. These scales provide important tools for educators who plan to implement new teaching strategies, and are

interested in more than just the performance outcomes that result from those strategies. The follow-up study will provide detailed connections among the scales and student performance.

Methods

Participants

Quantitative studies. Participants for the initial study were recruited from all first-semester general chemistry laboratory sections at a mid-sized Rocky Mountain region university during the fall of 2013. This course is required by several science and health majors, and represents the first of two courses in the general chemistry sequence. Additional participants for the cross-validation studies were recruited during the fall semester of 2014 at the same institution as well as at a second institution in the same US region.

Qualitative studies. Interviews were conducted with students from the target population (first-semester general chemistry students) to gather evidence for the response process validity of the modified items and scales (AERA, APA, NCME, 1999). In the fall of 2013, when the scales were given to lab sections, the last item asked students to indicate whether or not they would participate in a short interview regarding the survey. Students who indicated interest were contacted through the school e-mail address they provided on the survey. In the spring of 2014, additional participants were recruited via an announcement during lecture. Interested students volunteered by adding their name to a sign-up sheet passed out and collected by one of the authors.

In accordance with Institutional Review Board (IRB) policy, students in both the quantitative and qualitative studies were informed that their participation had no impact on their course grade and that they would be volunteering for a research study regarding their academic motivation. Standard university policies for confidentiality and data handling were utilized throughout the study. Participation in the studies was voluntary and no incentives were given to students for participating.

Data Collection

Quantitative data for the initial study were collected in all laboratory sections of the course at two time points. At the pre-semester data collection (time 1), only students who were enrolled in a lab section and were present in lab during the 1st week of the semester could be included in the study. At the post-semester data collection (time 2), only students who were present in lab during the 13th week of the semester could be included in the study. At each time point, the teaching assistants for the labs gave a prescribed announcement regarding the purpose of the study and instructions for completing the survey. Following that, students were administered a packet containing the survey with all scales and a set of demographic items (see Appendix C for survey and demographic items). Each student was asked to provide an identifier, so that time 2 responses could be matched to that specific student. Students were given approximately 15 minutes to complete the items on the survey and six demographic items. All students were required to take the survey, but were informed that their data would only be used if they signed the consent form. The consent form covered

both the time 1 and time 2 data collections and was only offered during time 1; thus, scores from the time 2 administration were not retained unless they could be matched to consent forms from the time 1 administration. In addition to individual time 1 and time 2 analyses, a matched-pair sample was also used for analyses involving changes in students' scores across the semester. Hence, two different sample sizes were used in the study. The total number of complete data sets from the time 1 administration was 373. The total number of complete data sets from time 2 (and hence the matched-data set pairs) was 294. All available and complete sets of scores from the start of the semester were retained, as we wanted to test the functionality of the scales with the entire incoming population, regardless of their future trajectory in the course. Data for the cross-validation studies were collected following the same protocols as noted above, with two exceptions: demographic data were not collected for the entire sample, and time 2 data were only collected at the main institution.

Scales

Preliminary wording changes. In order to appropriately assess chemistry students on the three latent traits (self-efficacy, interest, and effort beliefs) being measured, we made minor wording changes and adjustments to the measurement scales where needed. These preliminary modifications are described below for the original scales. All quantitative and qualitative study participants were given these modified scales (hereafter referred to as "scales"). Changes made to the scales after administration to the students and interviews

were data-driven. The resulting scales following these data-driven revisions hereafter are referred to as “revised” scales.

Chemistry self-efficacy. The self-efficacy scale was taken from the College Chemistry Self-Efficacy Scale (CCSS; (Uzuntiryaki & Aydin, 2009). These items are designed to measure a student’s perception of his or her ability to complete a given task in a chemistry course. The original instrument (21 items) has three subscales of chemistry self-efficacy: self-efficacy for cognitive skills (12 items), self-efficacy for psychomotor skills (5 items), and self-efficacy for everyday application (4 items). The original items are on a 9-point Likert-type scale ranging from “very poorly” to “very well.” Select items from the self-efficacy for cognitive skills subscale were evaluated for this study. Uzuntiryaki and Aydin (2009) reported a Cronbach’s alpha estimate of 0.92 for this subscale. These items are intended to measure students’ belief in their ability to work through intellectual operations in chemistry (Uzuntiryaki & Aydin, 2009). Our main interest in students’ self-efficacy is related to chemistry problems encountered during the lecture portion of class; thus, items related to the laboratory or the nature of science were excluded (see Appendix C for items). An example item that was excluded is: “How well can you write a lab report summarizing main findings?” Example items that were retained include: “To what extent can you explain chemical laws and theories” and “How well can you read the formulas of elements and compounds?” In the original instrument developed by Uzuntiryaki and Aydin (2009), nine numerical choices were given, but only five delineated categorical choices, ranging from “very poorly” to “very well”, were placed above

the numbers. The student then, is left with multiple numerical choices per category. The meaning of the difference between the two numerical choices is therefore lost. Hence, we argue that for clarity, if only five categories are given, then only five numerical choices are necessary. As there was no compelling reason to retain the nine options, and to allow for electronic scoring, we changed the nine-point Likert-type scale used in the original instrument to a five-point Likert-type scale. Since we are testing the internal structure of the scale, a CFA result consistent with the CFA from the original authors (Uzuntiryaki & Aydin, 2009) will provide favorable support for the condensed number of response options. In addition, there is evidence to support that changing scale length does little to affect the distribution about the mean, skewness, or kurtosis (Dawes, 2008).

Initial and maintained interest. The original initial interest and maintained interest scales were adapted from a survey developed by Harackiewicz et al. (2008). The initial interest items were designed to measure a student's interest in psychology at the beginning of an introductory undergraduate psychology course. The maintained interest items were given to students at week 13 of the semester and were designed to measure the "hold" component of situational interest. Exploratory and confirmatory factor analysis performed with these items confirmed the distinction between the "catch" and "hold" components of situational interest (Linnenbrink et al. 2010). The original scales for initial interest and maintained interest have seven and nine items, respectively. Both original scales are measured on a seven-point Likert-type

scale ranging from “not at all true of me” to “very true of me.” The Cronbach’s alpha estimates ($\alpha = 0.90$ for the initial interest scale and $\alpha = 0.95$ for the maintained interest scale) for scores based on a sample of 1,265 college students in an introductory psychology class were deemed acceptable (Harackiewicz *et al.*, 2008). We modified the wording of these items slightly to fit the context of a chemistry course, mostly by just replacing the word “psychology” with the word “chemistry.” For example, the item, “I am really looking forward to learning more about psychology” was changed to “I am really looking forward to learning more about chemistry.” As the authors provided no rationale for retaining the 7-point scale, and to keep the number of response options consistent across all measures, we adjusted the scale to 5-points. Lastly, the responses on the scale were changed to from “true of me” statements to “agree” statements (see Appendix C for items) as agree-type response options better fit the wording of the items.

Effort beliefs. The original items for the effort beliefs scale were developed by Sorich and Dweck (1997) and first used in Blackwell’s (2002) unpublished doctoral dissertation study, which involved seventh grade students. The nine-item effort beliefs scale was designed to measure the degree to which students believe their effort will lead to positive outcomes. These items were then adapted for use in a motivational study involving ninth grade math students (Jones *et al.*, 2012). The effort beliefs scale used by Jones *et al.* (2012) consisted of nine items measured by a six-point Likert-type scale ranging from “strongly disagree” to “strongly agree.” Jones *et al.* (2012) found the Cronbach’s

alpha ($\alpha = .77$) estimate for a sample of 163 ninth-grade math students acceptable. We used the exact wording of each item (based on Jones et al. (2012) version), except for substituting the word “chemistry” for the word “math.” The scale consists of four positive items (“If a chemistry assignment is hard, I’ll probably learn a lot doing it”), and five negative items (“To tell the truth, when I work hard at chemistry, it makes me feel like I’m not very smart”). In addition, we adjusted the scale range from a 6-point to a 5-point Likert scale.

Qualitative

Interview protocol. All interviews took place in a private interview room to ensure both participant confidentiality and audio quality. Prior to starting the interview, each participant was informed about the purpose of the study, the interview procedure, and the protocols for confidentiality. Following that, the participants signed a consent form approved by the IRB. Since students interviewed in the fall 2013 had already completed the scales at the start of the semester, they were provided a copy of their original answer choices for each scale. Students who were interviewed in the spring of 2014 were asked to complete the scales prior to the start of the interview. All interviews were audio recorded. A verbal probing interview approach was used, whereby students were asked to read each item out-loud, explain the reasoning behind the answer choice they made, and comment on the readability of the items (Knafl *et al.*, 2007). If a student’s reasoning did not match their answer choice or was unclear to the researcher, additional probing questions were asked to clarify their interpretation of the item. This methodology is important in establishing the

response process validity (Arjoon *et al.*, 2013) for the modified items and response scales, ensuring proper readability and consistency between students' answer choices and reasoning among the target population (Barbera & VandenPlas, 2011).

Quantitative Data Analysis

Descriptives. Descriptive statistics were analyzed on all data to check for skew and kurtosis as well as to determine means and standard deviations. In line with what is commonly accepted, we considered acceptable skew and kurtosis values as falling in the range of -1 to +1 (Huck, 2012). Reliability estimates for internal consistency (Cronbach's α) were calculated for each scale as well. Cronbach's α is an estimate of the internal consistency in the responses and should be reported with respect to each scale (AERA, APA, NCME, 1999). A value of 0.70 is considered acceptable for classroom multiple-choice tests and rating scales (Murphy & Davidshofer, 2005). Statistical Package for the Social Sciences (SPSS) 20.0 software was used for these analyses.

Time 1 and time 2 measurements. Analysis of variance (ANOVA) tests were performed on each scale at both time points separately to determine if any significant differences existed among students' scores based on major choice. Paired samples *t*-tests were conducted to determine if any changes were significant in students' scores across the semester. All item scores from each scale were aggregated to produce a scale score. In congruence with what has been commonly reported in the field of chemical education and among authors of the scales used in this paper, a mean score for each scale, based on the raw

item scores, was produced (Blackwell, 2002; Dalgety & Salter, 2002; Zusho *et al.*, 2003; Lewis *et al.*, 2009; Jones *et al.*, 2012). The mean score aggregation method was used as each scale contains a different number of items. Therefore, this method allows for consistent interpretation across scales and will lead to less variance when utilizing scale scores in future path analyses (Kline, 2011). To further assess if these changes were different by major, mixed-between-within ANOVA tests were employed for the interest and self-efficacy scales. As we do not have any theoretical underpinnings or prior studies to support the existence of differences in effort beliefs by academic major, we cannot use this type of comparison to provide supporting evidence of validity for the effort beliefs scale. Only students who took the survey at both time points in the initial study were included in this analysis ($n = 294$). All tests were evaluated at $p < 0.05$.

Confirmatory factor analysis. Confirmatory factor analysis (CFA) is a powerful tool for assessing how well a proposed model fits a set of measured variables. To date, a few studies in chemical education have used CFA during scale development and validation (Uzuntiryaki & Aydin, 2009; Xu & Lewis, 2011; Raker *et al.*, 2013). Each scale (i.e., self-efficacy, initial interest, and effort beliefs) was considered a latent variable (or factor) and each item, an indicator for its respective scale. To identify any problematic items that should be considered for deletion and ascertain the fit of each indicator to the appropriate latent variable, a one-factor CFA was conducted for each scale. One item per latent variable was set to unity. Only complete data sets were included in the analyses; thus, list-wise deletion was used for any missing data. All CFAs were

performed using LISREL version 8.80 (Jöreskog & Sörbom, 2006). Analyses were based on the robust maximum likelihood (RML) estimator, as the data were treated as ordinal and were non-normal. The commonly used maximum likelihood (ML) estimator is not appropriate for these analyses because the data must be continuous and normal for the ML estimator to be unbiased with respect to fit indices, parameter estimates, and standard errors (Finney & DiStefano, 2006).

Global fit of each one-factor model was analyzed based on several indices including: Satorra-Bentler (SB) scaled chi-square (Satorra & Bentler, 1994), root mean squared error of approximation (RMSEA); (Steiger, 1990), non-normed fit index (TLI); (Tucker & Lewis, 1973), standardized root mean squared residual (SRMR) , and comparative fit index (CFI); (Bentler, 1990). The SB scaled chi-square is a test for exact model fit, where the population covariances are fully reproduced by the hypothesized model. A non-significant result is desired and indicates that there are not significant discrepancies between the population covariances and those predicted by the model (Kline, 2011). However, the chi-square test is sensitive to sample size and will often produce a significant result for very small deviations in model fit. Thus, other descriptive test statistics are used to assess the fit of the model (Schermelleh-Engel, Moosbrugger, & Müller, 2003).

There are two types of fit indices commonly used in the literature to assess model fit: absolute and incremental. Absolute fit indices (i.e., chi-square, RMSEA, and SRMR) are estimates of how well an *a priori* model fits the data.

Incremental, or comparative fit indices (i.e., TLI and CFI) reflect improvement of model fit compared to a baseline model (Kline, 2011). The RMSEA can range from 0 to infinity and is a measure of the approximate model fit in the population (Steiger, 1990). Because exact fit of the model in the population is impractical, the RMSEA is a measure of “close fit,” and in general, values < 0.05 are considered good and those < 0.08 are considered reasonable (Browne and Cudeck, 1992). The SRMR value ranges from 0 to 1 and is a “badness of fit” measure based on the standardized fitted residuals. By standardizing the residuals, the scale of the variables is taken into account (Schermelleh-Engel *et al.*, 2003). Hu and Bentler (1995) suggested that an SRMR value of < 0.05 is indicative of good fit and < 0.10 is acceptable. The TLI and CFI both take into account the chi-square values of the proposed model and the null baseline model (Brown, 2006). The TLI and CFI values are normed and range from 0 to 1, with values ≥ 0.95 indicating good fit (Hu & Bentler, 1999). Only when several fit indices (both incremental and absolute) are considered together can the quality model fit be assessed with reasonable propriety (Brown, 2006). Based on what is commonly accepted in the literature, we used the following cut-off values as an evaluation of acceptable model fit beyond the chi-square test statistic: $RMSEA \leq 0.05$, $SRMR \leq 0.10$, TLI and $CFI \geq 0.95$ (Hu & Bentler, 1999; Hooper *et al.*, 2008).

Component model fit was evaluated based on statistical significance ($p < 0.05$) and reasonable parameter estimates. In addition, modification indices were considered when significant. The modification index (MI) is represented as a one

degree of freedom chi-square statistic that estimates the difference between two nested models. Modification indices are parameter-specific and reflect the approximate decrease in the model chi-square statistic when the fixed parameter is allowed to be freely estimated (Brown, 2006). With regard to CFAs, MI values are most commonly associated with correlated residuals between two indicator variables or between an indicator and two factors in the model. Thus, the higher the MI, the more likely it is that a particular indicator either is redundant or belongs with another factor. Modification indices are part of the evidence used to assess whether an indicator (item) should be dropped, aggregated with another indicator, or relocated to a different factor (latent trait). When the MI value exceeds the critical χ^2 value ($\alpha = 0.05$) using the degrees of freedom (*df*) from the model, then the appropriate modification should be considered (Hancock, 1999).

The one-factor CFAs were conducted on data collected at each time point. At time 1, all complete data sets were included for those students who consented to the study ($n = 373$). At time 2, only matched data sets were used in the analysis ($n = 294$). Although more than 294 students participated at the end of the semester, consent forms were only issued at the beginning of the semester, precluding the use of data from students who might have been absent at the first data collection or added the class late.

Qualitative Data Analysis

All interviews were transcribed and coded for significant statements and emergent themes, based on each item and its corresponding scale (Creswell, 2013a). The strategy for coding centered on readability and the degree of

consistency among participants' interpretations of the items. If students repeatedly report dissimilar interpretations of an item, then there can be no consensus on what the score of that item really means. This is problematic, as it negatively affects the validity of the scale and the inferences that can be drawn from the scores.

Results

Interview Results

A total of nine interviews (2 males and 7 females) were conducted in the fall of 2013 and five additional interviews (4 males and 1 female) in the spring of 2014. We felt that this number of interviews was sufficient to reach consensus on the items, especially given that we were not designing them from scratch. The students who participated in the fall 2013 interviews were asked to comment on the initial interest, self-efficacy, and effort beliefs scales. The interviews were conducted during the middle of the semester; therefore, students had not yet been given the maintained interest scale. For this reason, and to solidify the results from the self-efficacy scale, the interviews in the spring of 2014 covered the maintained interest and self-efficacy items. The effort beliefs scale and initial interest scales were not included in the spring interviews because we observed consistent responses from participants regarding the meaning and interpretability of items during the fall interviews.

In-depth student interviews were conducted using the full versions of the scales. The results from the interviews were used in conjunction with quantitative

results from the CFAs as support to flag any items that should be considered for modification or removal.

Readability and interpretation. Overall, the items in all scales showed good readability during the interviews. Participants read most of the items without stumbling and seemed to have a good grasp on the flow of each statement. However, there was one item in the effort beliefs scale that failed to show adequate readability and was confusing for many of the participants. Item 7 (see Appendix C for items) reads, “If you don’t work hard in chemistry and put in a lot of effort, you won’t do well”. During the interview, many participants had to read the item at least twice before explaining their reasoning for the answer they chose. Several participants regarded the item as “confusing”. Below is an example of how one student had to double back on his answer choice.

Oh, it was a little confusing, I guess, yea. I would have said – I kind of connected it with the other ones and just said strongly agree. But I guess – but now I would say I disagree because...Oh, wait, let me re-phrase that. Yeah, it’s a little confusing, the wording. I strongly agree with that too.

This student started off with an affirmative response, then he switched his answer to a negative response, before returning to his original choice. The confusion over the wording of this item was consistent throughout the interviews.

In addition to readability, we were interested in how students attributed meaning to the items. Toward this end, participants were asked to explain their reasoning behind the answer choice they made. Most participants gave plausible reasons for the answers they chose, and provided rational explanations. However, there were two items from the self-efficacy scale (Items 3 and 8) for

which the agreement of the meaning differed among several students. Item 3 (“How well can you describe the structure of an atom?”) was problematic because there were differing opinions about what it meant to “describe the structure of an atom.” Some students reported describing the structure of an atom simply meant knowing “the positions of things and charges”. Others reported that “interactions” and valence shell theory were part of the description. We observed, in several cases, that this item could be interpreted to varying degrees of depth and understanding. Item 8 (“How well can you solve chemistry problems?”) was also problematic for a similar reason. Participants regarded this item as “broad”, “vague”, and “depending on the problem”. Clearly, there are many types of problems students encounter in first-semester general chemistry. Diffuse tasks such as these can lead to problems when a student tries to self-appraise their ability to complete the task (Bandura, 1986). Additionally, item 8 is somewhat redundant in that every item that precedes it represents some type of chemistry problem.

Feeling-related interest versus value-related interest. Individual interest is conceptualized as having both feeling-related (emotional arousal) and value-related (importance/utility) components (Schiefele, 1999). The scale used in this study was designed to measure initial interest, thus items to measure both components were incorporated into the scale. However, in the original study from which the items were adapted, initial interest was presented as a single factor (Harackiewicz *et al.*, 2008). In a later study on situational interest, a similar set of items were grouped into two factors of interest: feeling-related and value-related

(Linnenbrink-Garcia *et al.*, 2010). As the factor structure is an important part of a scale and the validity of the gathered data, we were concerned with how students responded to feeling-related versus value-related items. If similar reasoning were given for all interest items, then the qualitative evidence to split the scale into two factors would be missing. If, however, there was a clear demarcation between reasons used for answers to feeling-related versus value-related items, then two factors might be a more valid interpretation of the scale.

We found that participants during the fall and spring interviews used dissimilar language when describing their reasons for answers to feeling-related items versus value-related items. Examples of feeling-related items from the initial interest scale are: "I am fascinated by chemistry" and "I chose to take general chemistry because I'm really interested in the topic". Participants cited reasons for choosing their answers by using words and phrases such as: "I'm naturally gifted", "I connect with the material", "It excites me", "Interested", "Fascinated". These words are evoked from "feelings of involvement, stimulation, and enjoyment" toward the topic of chemistry, and is exactly the type of interest that is characterized by feeling-related items (Schiefele, 1999). On the other hand, participants explaining their answer choices for value-related items used entirely different language. Examples of value-related items from the initial interest scale are: "I think what we will study in general chemistry will be important for me to know" and "I think the field of chemistry is an important discipline". Participants commented on their answer choices by using phrases such as: "I'm going to be building off this", "it will obviously be important in my

field”, “this class will...help with future chemistry classes”, “chemistry is...everything around us”. Schiefele (1999) describes value-related interest as being directed toward something that is personally significant and important to the individual. The statements made by the participants during the interviews regarding value-related items were indicative of a personal significance and importance, as opposed to feelings of excitement or enjoyment. Based on the overwhelming difference we found in how participants described their interest using feeling-related items versus value-related items, the scale was tested as both a 1 and 2-factor model.

Data Screening and Descriptive Statistics

Prior to analysis, all data sets were screened for careless responses from students (i.e., students selecting all of one response option). Only one case was found that exhibited this pattern. Missing item-level data were also screened for patterns. The only consistently missed item was the last item on the list. Of 37 total cases that had missing data, 17 of them failed to respond to the last item. As list-wise deletion was implemented, all cases with missing data were removed from each data set prior to analysis.

Mean, standard deviation, skew and kurtosis were evaluated for each item on each scale (see Table C1). Most items had skew and kurtosis values that were within acceptable ranges to be considered normal (-1 to +1) (Huck, 2012). However, some items were outside of this range with negative skew values down to ~ -1.5 and kurtosis values up to 3.5. Due to these deviations from normality, the robust maximum likelihood (RML) estimator was used in all CFA runs. The

RML estimator utilizes the Satorra-Bentler scaled chi-square statistic, and is robust with respect to non-normal data (Chou *et al.*, 1991).

Demographics. Demographic data were collected from all participants in the initial study during the fall of 2013. In the fall of 2014, a cross-validation study was conducted ($n = 1160$), but demographic data were only collected from a subsample of these participants ($n = 175$), those for whom both pre- and post-semester data were gathered. Of the participants in the initial study ($n = 373$, pre-semester), most were female (67%). Nearly half (46%) were non-science majors (nursing, sports and exercise science, statistics, earth science), 32% were other science majors (biology, physics, or mathematics), 20% were chemistry majors, and 2% were undeclared. In this sample, the majority of students were first-year (60%) and second-year (13%) university students. Most took a chemistry course in secondary school (89%). Similar demographic breakdowns from the initial study were observed for the matched-pair sample ($n = 294$), with all categories within 1 percentage point of the reported statistics. Demographic data from the cross-validation study ($n = 175$) showed a similar breakdown. Most of the students were female (73%), and most reported taking chemistry in secondary school (93%). Non-science majors made up the bulk of the sample (62%), followed by other science (24%), and chemistry (13%). Nearly 80% of the sample was first and second-year university students.

Reliability analysis. The internal consistency estimates (Cronbach's α) for each scale were acceptable to high, ranging from 0.75 to 0.88 (see Tables 1 and 2). As revisions were made to the scales, based on qualitative data, fit

indices, and factor loadings from the single factor CFAs, the alpha values dropped for two of the scales (self-efficacy and initial interest), and increased for the effort beliefs scale. Despite this, all alpha values remained acceptable to high (0.77 to 0.89).

Time 1 CFAs. Each scale was evaluated using a single-factor CFA using a sample of first-semester general chemistry students at the start of the semester ($n = 373$). The goals of the analyses were to substantiate each scales structure and to seek possibilities to shorten the scales. From the quantitative side, revisions to the model (e.g., dropping items) were guided by the fit indices as well as the modification indexes for each scale. We considered parallel qualitative evidence together with the CFA results before making decisions about model revisions. Table 1 shows the values of the fit indices for each time 1 CFA (χ^2 , RMSEA, NFI, CFI, and SRMR) before and after revision of the model.

Table 1. CFA fit indices and reliability estimates of preliminary and revised scales at time 1 for the initial sample

Scale	# of items	χ^2 value	df^a	p -value	RMSEA	TLI	CFI	SRMR	α^b
Initial interest	7	158.81	14	< 0.001	0.17	0.93	0.95	0.09	0.88
Initial feeling	4								0.90
Initial value (revised)	3	23.84	13	0.033	0.05	0.99	1.00	0.03	0.79
Effort beliefs	9	148.25	27	< 0.001	0.11	0.92	0.93	0.10	0.75
Effort beliefs (revised)	6	16.27	9	0.26	0.05	0.99	0.99	0.05	0.77
Self-efficacy	8	96.49	20	< 0.001	0.10	0.97	0.98	0.06	0.89
Self-efficacy (revised)	6	43.32	9	< 0.001	0.10	0.97	0.98	0.05	0.85

$n = 373$, ^aDegrees of freedom (df) are based on RML estimator ^bCronbach's alpha χ^2 – likelihood ratio test, RMSEA – root mean squared error of approximation, TLI – Tucker-Lewis index, CFI – comparative fit index, SRMR – standardized root mean squared residual

Initial interest. The model for initial interest was based on seven indicators for the 1-factor solution (see Figure C1). The standardized factor loadings for all seven indicators were significant ($p < 0.05$). However, after analyzing the fit and modification indices, it was clear that the model did not adequately fit the data. The global fit index, SB-scaled chi-square test, indicated inadequate fit of the model to the data $\chi^2 (14, n = 373) = 158.81, p < 0.001$. However, a significant chi-square test is very common with large sample sizes. Component fit indices, RMSEA value (0.17) and SRMR (0.094), also suggested inadequate fit (see Table 1). In addition, items 5, 6, and 7 displayed markedly lower standardized

factor loadings than items 1 – 4. Based on these results, and the qualitative evidence suggesting students use different language when describing feeling-related versus value-related interest, the scale was split into two factors. A second CFA was run with the scale split into feeling-related and value-related factors. All of the reported fit indices improved for the 2-factor model, meeting the acceptable cut-off values. Although the SB-scaled chi-square test remained significant, $\chi^2 (13, n = 373) = 23.84, p < 0.033$, the improvement of other fit indices following revision, suggested reasonable fit.

Effort beliefs. The 1-factor model for effort beliefs was composed of nine indicators, five of which were negatively worded and were reverse-coded for analysis (see Figure C2). The RMSEA (0.11), TLI (0.92), and CFI (0.93), as well as low standardized factor loadings (< 0.40) for some items, suggested inadequate fit of the model. Three items (1, 7, and 9) were dropped based on low factor loadings (< 0.40). Moreover, items 7 and 9 had large modification indices with other items in the scale, suggesting correlated error among those items. Once these three items were removed, a second 1-factor CFA was run and all of the reported fit indices improved to be within the appropriate ranges considered acceptable. In addition, the SB-scaled chi-square statistic, $\chi^2 (9, n = 373) = 162.7, p = 0.061$, was not statistically significant. Taken together, these results suggest the revised model fits the data well.

Self-Efficacy. The 1-factor model for self-efficacy was composed of eight indicators (see Figure C3). The RMSEA (0.10) and SRMR (0.06) values, as well as large modification indices for several items, suggested poor fit of the model.

Items 2, 3 and 8 all had high modification indices with at least two other items. Additionally, items 3 and 8 were found to be problematic in the student interviews due to ambiguity in the meaning that students attributed to them. Hence, both items were removed from the model. Item 2 was left in the model as a high modification index should not be the only criteria for removing an indicator from a model and no other quantitative or qualitative results supported removal. A second 1-factor CFA was run with the revised scale, but the fit indices did not suggest improved model fit. The removal of these items was neither an improvement nor a detriment to the model fit. In spite of this, we chose to retain the revised scale. We feel that the qualitative data and high MI values are sufficient reasons to justify removing these items, and therefore shortening, the scale. Although the original authors of this scale had a larger CFA model with two additional subscales, our CFA results for selected items from the self-efficacy for cognitive skills (SCS) subscale are consistent with those from the authors (Uzuntiryaki & Aydin, 2009). The only exception to this is the RMSEA value from our model, which was slightly inflated (0.10) compared to the original authors' model (0.08).

Time 2 CFAs. The self-efficacy and effort beliefs scales consisted of the same items from time 1 to time 2. CFAs, using the revised models from time 1, were run on these two scales to confirm the revised scale structures and functionality of items (see Figures C5 and C6). We found that the two revised models fit the time 2 data adequately (see Table 2). The items for maintained

interest were not identical to those of initial interest, thus a 1-factor model was evaluated and subsequent revisions made.

Maintained interest. The 1-factor model for maintained interest was based on eight indicators (see Figure C4). As with the initial interest scale, the maintained interest items were composed of both feeling-related and value-related interest. One item on the maintained interest scale was negatively worded and was reverse coded prior to analysis. When the 1-factor model was run, the fit was very poor, $\chi^2 (20, n = 294) = 222.34, p < 0.001$, RMSEA (0.19), TLI (0.92), CFI (0.94), and SRMR (0.12). All reported global and component fit indices were outside of the acceptable ranges. When the model was split into two factors (feeling-related and value-related), the fit improved dramatically. The SB-scaled chi-square statistic was not significant ($p = 0.26$), suggesting adequate global model fit. Additionally, all component fit indices improved, falling within good to acceptable ranges. Standardized factor loadings for the indicators were significant ($p < 0.05$) with both models.

Table 2. CFA fit indices and reliability estimates of preliminary and revised scales at time 2 for the initial sample

Scale	# of items	χ^2 value	df^a	p -value	RMSEA	TLI	CFI	SRMR	α^b
Maintained interest	8	222.34	20	< 0.001	0.19	0.92	0.94	0.12	0.91
Maintained feeling	4	22.49	19	0.26	0.03	1.00	1.00	0.04	0.92
Maintained value (revised)	4								0.87
Effort beliefs (revised)	6	15.49	9	0.08	0.05	0.99	0.99	0.03	0.82
Self-efficacy (revised)	6	36.03	9	< 0.001	0.10	0.97	0.98	0.04	0.87

Cross-validation of revised scales. Supporting evidence for the structural validity of the revised scales is provided by evaluating the model fit with alternate samples from the same target population (Kline, 2011). It is advised that anytime a model is revised, the revised model be cross-validated with an independent sample. Samples for cross-validation studies could come from the original data set, if the sample size is large enough, or from a completely separate data collection (Brown, 2006, p. 124). As the initial data from fall 2013 was used to make revisions to each scale, cross-validation samples were collected in fall 2014 and used to further validate the revised scales. The first sample (cross-validation 1) was collected at the same institution as the initial data set, a second sample (cross-validation 2) was collected at a different institution. Due to administration constraints, post-semester data were not collected from the second sample; therefore, the maintained interest items were not cross-validated with this population. A comparison of fit indices and reliability estimates for all samples is presented in Table 3.

Table 3. CFA fit indices and reliability estimates for initial and cross-validation samples

Fit Index	Revised scales (n = 373)				Cross-validation 1 (n = 432)				Cross-validation 2 (n = 728)		
	II	MI	EB	SE	II	MI ^c	EB	SE	II	EB	SE
χ^2	23.8	22.5	16.3	31.3	49.0	18.3	60.0	62.7	80.4	48.6	66.8
df ^a	13	19	9	9	13	19	9	9	13	9	9
p-value	0.03	0.26	0.26	0.00	0.00	0.50	0.00	0.00	0.00	0.00	0.00
RMSEA	0.05	0.03	0.05	0.08	0.08	0.00	0.11	0.12	0.08	0.08	0.09
TLI	0.99	1.00	0.99	0.98	0.99	1.00	0.96	0.96	0.98	0.97	0.96
CFI	1.00	1.00	0.99	0.99	0.99	1.00	0.98	0.98	0.99	0.98	0.98
SRMR	0.03	0.04	0.05	0.04	0.05	0.04	0.08	0.04	0.04	0.06	0.05
α^b	0.90	0.92	0.77	0.85	0.91	0.94	0.84	0.88	0.90	0.75	0.82
	0.79	0.87			0.90	0.86			0.84		

^aDegrees of freedom (df) are based on RML estimator ^bCronbach's alpha ^cPost-semester n=175, χ^2 – likelihood ratio test, RMSEA – root mean squared error of approximation, TLI – Tucker-Lewis index, CFI – comparative fit index, SRMR – standardized root mean squared residual

Fit indices and alpha values from the cross-validation samples confirm the validity and reliability of the revised scales. All scales had consistently high Cronbach's alpha values across all samples, indicating similar reliability for each administration. With only a few exceptions, the fit indices were within the range of acceptable values. In the cross-validation 1 sample, the effort beliefs scale produced SRMR (0.08) and RMSEA (0.11) values above the standard cut-offs (≤ 0.06 and ≤ 0.08). However, Hu and Bentler report that acceptable SRMR values can be as high as 0.09 with CFI and TLI values > 0.95 (Hu and Bentler, 1999). Therefore, with only the RMSEA value being out of range, the revised scale model has acceptable fit to the data from this sample. The self-efficacy scale produced inflated RMSEA values (0.12 and 0.09) in both cross-validation samples; however, the other indices were well within the acceptable ranges. Therefore, the revised self-efficacy scale is deemed to have acceptable fit to both populations. As both additional data sets were larger than the original data set, it is expected that the SB-scaled chi-squared values would be significant as the

chi-square test is highly dependent on sample size (Brown, 2006). The only exception to this was the maintained interest scale, as this post-semester data set only contained 175 students. Taken together, the consistent fit indices and alpha values across all samples provide supporting evidence for the validity and reliability of the revised scales across three different samples from two different institutions.

Pre- and post-semester comparisons. The comparison of pre-semester (time 1) with post-semester (time 2) scores was conducted using only the matched sample data sets from the initial sample ($n = 294$) and the cross-validation study sub-sample ($n = 175$). To check for data patterns from those students who did not have time 2 responses, demographic items and item-level means from time 1 were compared between those with and without time 2 data. The frequencies of responses to all demographic items and item-level means appeared very similar, indicating that the two groups of students were likely from the same population. The initial interest and maintained interest scales address related traits, however, due to wording differences each scale consisted of different items from time 1 to time 2. Hence, the mean scores from the scales (Table 4) cannot be directly compared with a t -test. However, the scales could be used to compare sub-groups of students (e.g., major choice, ethnicity, or gender) based on how their interest changed relative to one another. In a future study, we will be using these scales as part of a path analysis to investigate the connections between these motivational factors and student performance. There was a significant drop in effort beliefs ($M_1 = 3.95$, $M_2 = 3.77$, see Table 5) among

all students, $t(293) = 4.35$, $p < 0.001$. However, the effect size ($d = 0.29$) was small (Cohen, 1992). This trend was observed for the cross-validation sub-sample as well ($M_1 = 4.03$, $M_2 = 3.91$) with a significant drop in effort beliefs scores, $t(174) = 2.30$, $p = 0.02$. The effect size ($d = 0.18$) of this difference was also small. Of the three scales, the most change across the semester and variation by group in students' scores was observed for the self-efficacy scale.

Table 4. Mean values of scores for initial and maintained interest at times 1 and 2 for all students from the initial sample

Scale	Mean (SD) time 1	Mean (SD) time 2
Initial interest (feeling)	3.46 (0.84)	-
Initial interest (value)	4.06 (0.67)	-
Maintained interest (feeling)	-	3.23 (0.90)
Maintained interest (value)	-	3.61 (0.83)

$n = 294$. Note: Interest scales were different from time 1 to time 2 and cannot be directly compared.

Self-efficacy overall. Self-efficacy is one's self-appraisal of ability to complete a task. Our measure of chemistry self-efficacy included tasks that would be commonly encountered in a first-semester chemistry class, such as: explaining the structure of an atom, or choosing an appropriate formula to solve a problem. As instructors, we would expect our students to improve upon these tasks during the course of a semester and, we would expect their self-appraisals of ability to improve as well.

Our results suggest that students' chemistry self-efficacy increased across the semester for both the initial sample and cross-validation sub-sample (see Table 5). Self-efficacy scores were based on a mean composite score of the revised scale (6 items). The mean difference in scores for the initial study ($M_1 =$

3.29, $M_2 = 3.60$) for all students across the semester was significant, $t(293) = 8.23$, $p < 0.001$). The effect size ($d = 0.50$) for this comparison was medium (Cohen, 1992). Data from the cross-validation study sub-sample showed a similar trend ($M_1 = 2.87$, $M_2 = 3.69$), with the difference also being significant, $t(174) = 14.83$, $p < 0.001$). The effect size ($d = 1.19$) for this test was large (Cohen, 1992). On average, students from both samples felt more confident in their abilities to solve chemistry problems at the end of the semester than at the beginning of the semester.

While our results suggest an increase in self-efficacy for our overall sample, this trend might not hold for all students in the sample. For example, Villafane *et al.* (2014) reported differing trajectories in chemistry self-efficacy based upon ethnic group. Similar trends were observed by Zusho *et al.* (2003) with regard to performance in chemistry. They reported that the self-efficacy of students who were “low achievers” in chemistry dropped sharply across a semester, and those who were “average achievers” dropped slightly. In contrast, students who were “high achievers” reported higher self-efficacy at the end of the semester than at the start. Collectively, these two studies demonstrate that students’ self-efficacy trends across a semester depend on several factors, some of which may change during the semester. Hence, instructors should be aware and expectant of such differing trends in self-efficacy among their students, especially when evaluating the effectiveness of a novel approach to instruction.

Table 5. Mean values of scores for self-efficacy and effort beliefs at times 1 and 2 for all students from both the initial sample and cross-validation sub-sample

Study	Scale	Mean (SD) time 1	Mean (SD) time 2	Mean difference ^a (effect size ^b)	p-value
Initial (<i>n</i> = 294)	Effort beliefs	3.95 (0.57)	3.77 (0.68)	-0.18 (0.29)	< 0.001
	Self-efficacy	3.29 (0.60)	3.60 (0.65)	0.31 (0.50)	< 0.001
Cross- validation	Effort beliefs	4.03 (0.57)	3.91 (0.74)	-0.13 (0.18)	0.022
	Self-efficacy	2.87 (0.73)	3.69 (0.64)	0.71 (1.19)	< 0.001

^aBased on paired samples *t*-test ^bEffect size represented by *Cohen's d* – small (0.20), medium (0.50), large (0.80) (Cohen, 1992).

Self-efficacy and interest by major. We were interested in whether declared chemistry majors differed from those in other majors on self-efficacy and interest. We expected chemistry majors to score differently than other majors based on the nature of interest and self-efficacy in academic choice (Lent *et al.*, 1994). Barbera *et al.* (2008) found that chemistry majors were more interested in chemistry than non-majors. To confirm this and test the notion that chemistry majors would also be more likely to report higher self-efficacy than non-majors, we performed ANOVA tests on both time 1 and time 2 data (see Table 6). All majors (chemistry, other science, non-science, and undeclared) were compared in the ANOVA test; only the post-hoc results on chemistry versus non-science majors are reported in Table 6. The overall ANOVA model for the pre-semester self-efficacy data was significant, $F(3,293) = 4.20, p = 0.006$. The assumption of homogeneity of variances was not violated, as indicated by a non-significant result of Levene's test ($p > 0.05$). Post-hoc analysis with the Tukey test showed that chemistry majors' self-efficacy ($M = 3.49, SD = 0.62$) was higher than non-

science majors' ($M = 3.17$, $SD = 0.59$). This difference was significant at $p < 0.01$. The overall ANOVA model for the post-semester self-efficacy data was not significant, indicating that chemistry majors did not differ from other majors at the end of the semester.

The overall models for the two components of initial interest (feeling- and value-related) were significant for the initial sample, $F(3,293) = 20.87$, $p < 0.001$, and $F(3, 293) = 7.12$, $p < 0.001$, respectively. The assumption of homogeneity of variances was not violated for any of the tests performed, as indicated by non-significant results using Levene's test ($p > 0.05$). Post-hoc analyses with Tukey tests revealed that chemistry majors reported higher feeling- ($M_1 = 4.14$, $SD_1 = 0.84$) and value-related interest ($M_2 = 4.38$, $SD_2 = 0.77$) than non-science majors ($M_1 = 3.19$, $SD_1 = 0.74$, $M_2 = 3.91$, $SD_2 = 0.62$) (see Table 6). The same trend was evident in the post-semester data with the two components of maintained interest being significant for the initial sample, $F(3,293) = 12.93$, $p < 0.001$, and $F(3,293) = 7.72$, $p < 0.001$, respectively. Post-hoc analyses revealed that chemistry majors' reported higher maintained feeling- ($M_1 = 3.73$, $SD_1 = 0.94$) and value-related interest ($M_2 = 3.95$, $SD_2 = 0.91$) than non-science majors ($M_1 = 2.93$, $SD_1 = 0.81$, $M_2 = 3.38$, $SD_2 = 0.79$). All differences reported were significant at $p < 0.01$. The corresponding effect sizes (Cohen's d) are considered medium to large (Cohen, 1992).

A sub-sample used in the cross-validation study was also evaluated for differences in self-efficacy and interest by major. Participants did not differ in self-efficacy, at either pre or post-semester. Participants did, however, report different

levels of interest (feeling and value) based on major at the start of the semester, $F(3,174) = 9.30, p < 0.001$, and $F(3, 174) = 4.59, p = 0.004$, respectively. The same was true for maintained interest (feeling and value) at the end of the semester, $F(3,174) = 6.53, p < 0.001$, and $F(3, 293) = 4.68, p = 0.004$, respectively. Based on our prior results reported above, we hypothesized that chemistry majors would report higher levels of interest than non-science majors. Hence, we performed planned contrasts to test this hypothesis. Results indicate that chemistry majors showed significantly more initial feeling-related interest, $t(171) = 4.56, p < .001$, and value-related interest, $t(171) = 3.66, p < .001$, than non-science majors. The same was true for maintained-feeling, $t(171) = 3.67, p < .001$, and maintained-value interests, $t(171) = 2.93, p = .004$ (see Table C3).

Table 6. Mean scores and differences between chemistry majors and non-science majors on interest and self-efficacy scales for the initial sample

Scale	Chemistry majors $n = 57$ Mean score (SD)	Non-science majors $n = 136$ Mean score (SD)	Mean difference ^a (effect size ^b)	p -value
Initial interest (feeling)	4.14 (0.84)	3.19 (0.74)	.95 (1.20)	< 0.001
Initial interest (value)	4.38 (0.77)	3.91 (0.62)	.47 (0.67)	< 0.001
Maintained interest (feeling)	3.73 (0.94)	2.93 (0.81)	.80 (0.91)	< 0.001
Maintained interest (value)	3.95 (0.91)	3.38 (0.79)	.57 (0.67)	< 0.001
Self-efficacy (time 1)	3.49 (0.62)	3.17 (0.59)	.32 (0.53)	0.005

^aBased on Tukey's post-hoc tests ^bEffect size represented by *Cohen's d* – small (0.20), medium (0.50), large (0.80) (Cohen, 1992).

These data support the findings from the study by Barbera et al. (2008) concerning personal interest among chemistry majors versus other majors. In addition, these results expand upon the findings reported by Uzuntiryaki and Ayden (2009) whereby “[chemistry] majors scored higher than non-majors [on self-efficacy for cognitive skills]; however, they did not appear significant.” Most importantly, these data demonstrate the ability of the modified items and revised scales to discriminate between populations of students who would be expected to score differently on self-efficacy and interest in chemistry.

Pre and post-semester comparisons of mean self-efficacy scores were analyzed by major (Table 7). These results indicate that all students with a declared major ($n = 287$) reported higher self-efficacy at the end of the semester than at the beginning of the semester, regardless of their major. All differences were significant at $p < 0.05$. Chemistry majors reported improved self-efficacy at the end of the semester ($M_2 = 3.70$) compared to the start of the semester ($M_1 = 3.18$), $t(56) = 5.39$, $p < .001$. The same was true for other science majors ($M_1 = 2.95$, $M_2 = 3.62$), $t(93) = 10.03$, $p < .001$; and non-science majors ($M_1 = 2.78$, $M_2 = 3.45$), $t(135) = 12.49$, $p < .001$. The effect sizes for the differences in self-efficacy from pre to post-semester were medium to large among all three groups of majors, ranging from $d = .71$ for chemistry majors to $d = 1.0$ for other science and non-science majors (Cohen, 1992). These results expand upon those reported in Table 5 by suggesting that students from all majors reported improved self-efficacy in chemistry after a semester of instruction.

Table 7. Mean values of self-efficacy scores by major at time 1 and time 2 from the initial sample

Major	Mean (SD) time 1	Mean (SD) time 2	Mean difference ^a (effect size ^b)	p-value
Chemistry	3.17 (0.69)	3.70 (0.79)	0.53 (0.71)	< 0.001
Other science	2.95 (0.66)	3.62 (0.61)	0.67 (1.00)	< 0.001
Non-science	2.78 (0.65)	3.44 (0.60)	0.66 (1.00)	< 0.001

$n = 287$, ^aBased on paired samples *t*-test ^bEffect size represented by *Cohen's d* – small (0.20), medium (0.50), large (0.80) (Cohen, 1992).

Discussion

Self-efficacy, interest, and effort beliefs are salient factors associated with academic motivation and are supported by a strong foundation of research (Weiner, 1985; Bandura, 1986; Renninger, 2000; Blackwell *et al.*, 2007). A chemistry-specific set of scales designed to measure interest, self-efficacy, and effort beliefs were administered to a sample of first-semester general chemistry students. The major goal of this study was to establish evidence of validity and reliability for scores from the scales such that they could be used in future studies regarding the impact of various teaching practices on these motivational factors and their relation to course performance. Demonstrating validity and reliability for data generated by a scale or instrument is paramount following any alterations to items or use in a new setting. Absence of such evidence renders the interpretations of scale scores invalid and can lead to misinformed decision-making (Arjoon *et al.*, 2013). In this study, we have presented evidence to support the internal consistency as well as the response process and structural validity of the modified items and revised scales.

Validity evidence based on response processes pertains to the agreement between the construct being measured and the actual processes respondents engage in when they generate an answer (AERA, APA, NCME, 1999). In the case of the interest scales, participants routinely described the meaning of the value-related items differently from the feeling-related items. These results, in conjunction with previous findings by Linnenbrink *et al.* (2010), led us to split the larger interest scales into two smaller subscales, which resulted in improved fit of the CFA models. Our revisions to the self-efficacy and effort beliefs scales were guided more by the way students interpreted the items. When participants assign varied interpretations to a particular item, the meaning of that item is obscured and integrity of the score and construct associated with that item is compromised. Items that were found to illicit ambiguous or frequently incongruent responses from participants were flagged as problematic. Those items that also demonstrated lack of fit or redundancy in the CFA models were removed from the corresponding scale. Overall, our interview results for the items retained in each scale suggest that students consistently and adequately understood the meaning of the items.

To demonstrate functionality based on the internal structure of the individual scales, 1-factor CFAs were conducted to examine the degree to which the data fit each hypothesized model. Confirmatory factor analysis allows the researcher to test whether a proposed grouping of items, and the scores associated with them, appropriately describe a latent variable (Brown, 2006). Our preliminary psychometric evaluation of each scale revealed that the model

fits were less than satisfactory. Revisions to each scale were informed by our qualitative studies with students from the target population as well as from the modification indices of the initial 1-factor CFA studies. Following revision of the effort beliefs and self-efficacy scales, and splitting of the interest scale into two factors, the model fit for each scale improved dramatically (see Tables 1 and 2). Nearly all fit indices improved to values considered acceptable for the CFAs conducted on data from both time points. The only exceptions to this were the RMSEA values for self-efficacy, and the chi-square values for initial interest and self-efficacy. As stated previously, the model chi-square test is a “badness of fit” index, where a significant result is not desired. Chi-square tests are sensitive to sample size and often, negligible deviations in fit produce significant results with large samples (Brown, 2006). The sample sizes ($n = 373$, $n = 294$) in our models would be considered large based on a recommended subject to indicator ratio of 1:10 (Bentler & Chou, 1987). Thus, it is not surprising that all of our preliminary models and two of our revised models failed to produce non-significant chi-square values. However, consistent with most studies involving factor analysis, we used approximate fit indices as alternative indicators of adequate model fit. The model describing the self-efficacy scale was the only one that remained problematic after revision, due to the significant chi-square value ($p < 0.05$) and high RMSEA value (0.10). Like the model chi-square test, the RMSEA value is considered a “badness of fit” index, where lower values are desired (Kline, 2011). Kline (2011) points out that values ≥ 0.10 might signal a “serious problem” with the model. Consequently, we urge readers to interpret our results for this model

with caution, but also to consider that RMSEA values tend to be inflated in models with a small number of indicators (Breivik and Olsson, 2001).

As the revised scales were derived from qualitative and quantitative results with a single population, additional evidence for the structural validity of each revised scale was established using two additional samples. These cross-validation studies were comprised of data collected from students in the same course, at the same institution, during the year following the initial study. The second sample was obtained from students in the same course at a different institution. Supporting evidence for the structural validity of the revised scales was provided by obtaining similar fit indices with both student samples (Kline, 2011). Across all three populations the fit indices remained consistent. With the exception of the RMSEA values for the effort beliefs scale from sample 1 and the self-efficacy scale from both samples, all other indices were within recommended ranges. As the cross-validation samples sizes were larger, all of the SB-scaled chi-squared statistics were significant, however, this is not seen as a threat to the structural validity as this test is highly dependent on sample size (Brown, 2006). The maintained interest scale did have a non-significant chi-square value, however, data for this one scale was derived from a much smaller ($n = 175$) post-semester population.

To further support the functionality of the interest and self-efficacy scales in a college chemistry setting, we evaluated the extent to which chemistry majors differed from other majors. Based on prior studies, we operated under the assumption that declared chemistry majors would have higher self-efficacy and

interest toward chemistry (Barbera *et al.*, 2008; Uzuntiryaki & Aydin, 2009). Our initial data show that chemistry majors began the semester with higher self-efficacy and interest than non-science majors. However, the gap in self-efficacy scores closed by the end of the semester, indicating no significant difference based on major. For the cross-validation sub-sample, the difference in interest scores between chemistry majors and non-science majors was retained, but there was no significant difference in self-efficacy. It is possible that the lower sample size ($n = 175$), and hence, lower number of chemistry majors ($n = 23$) compared to the initial sample could be a cause of this discrepancy. Results from the maintained interest scale suggest that chemistry majors continued to have higher levels of interest than non-science majors through the end of the semester. This is certainly plausible and expected, given that enduring interest in a particular subject area has been shown to predict major choice and number of courses taken in that subject area (Harackiewicz *et al.*, 2000; Harackiewicz *et al.*, 2008). Taken together, these results suggest that the self-efficacy and interest scales can discriminate between groups for whom it is plausible to expect differences in confidence and interest toward chemistry, providing further validity evidence of the scales (Standard 1.14, AERA, APA, NCME, 1999).

The internal consistency of scales is an estimate of reliability that relates to how well items within a scale describe the same construct (Henson, 2001). Cronbach's alpha is the most commonly reported value of internal consistency, and a cutoff value of 0.70 is often used to indicate moderate internal consistency among items used in classroom rating scales (Murphy & Davidshofer, 2005).

Following item modifications and scale revisions, all of our scales had reliability estimates ≥ 0.77 . These results suggest that the items belonging to each scale are consistent with other items in the same scale in describing the specific construct. Consistently high alpha values were also obtained in the cross-validation studies, further supporting the internal consistency of the revised scales with the target population.

Limitations

While the results from our psychometric evaluation of these revised scales suggest that they function well among general chemistry students, we acknowledge several limitations to the study. First, our sample size from time 1 to time 2 in the initial study dropped by 79 participants (21%). While the time 2 sample size ($n = 294$) remained large enough for factor analysis, the missing participants could represent an important subset of the population (e.g., students who dropped the course). However, as with any study involving multiple collections of data from a single sample, there is always a risk of attrition. While our cross-validation studies provide positive support for the generalizability of the revised scales, we encourage researchers from other institutions to use and further evaluate this set of scales, so that educators can have a more complete understanding of the psychometric properties and generalizability of these scales. Finally, we acknowledge that the meaning attributed to items in the self-efficacy scale may be different among students in the same population. We excluded two items from the scale (items 3 and 8) due to ambiguity and a lack of consensus on the meaning among interview participants. However, these were

items that had the most frequently incongruent responses. Participants in our sample did not necessarily assign the exact same meaning to the remaining items. For instance, item 4 reads, *How well can you describe the properties of elements by using the periodic table?* A student with a strong background in chemistry might, for example, interpret “properties of elements” as electronegativity, ionization energy, and bonding tendencies of elements. On the other hand, a student with a weak background in chemistry may view “properties of elements” as simply the number of protons, neutrons, and electrons in a given element. Thus, students with a strong chemistry background may, in fact, underestimate their ability because they have a deeper understanding of the theories and facets associated with certain tasks in chemistry. Students with less understanding of a chemistry task may inflate their appraisal of their ability due to an oversimplification of the task. In addition, students who are from non-English language backgrounds may interpret self-efficacy items differently than native English speakers (Lee & Fradd, 1998). Self-efficacy items involving tasks that are not completely objective and defined will always leave room for loose interpretations. However, items that are too narrowly focused and specific will lose generalizability and require the instrument to be long and arduous in order to cover the set of topics in a given course. On the other hand, Pajares (1996) cautions against using measures too general by stating that, “omnibus tests...transform self-efficacy into a generalized personality trait rather than a context-specific judgment.” The balance between specificity and generality with efficacy beliefs is a difficult aspect to fully resolve, a sentiment shared by other

researchers as well (Tschannen-Moran & Hoy, 2001). We feel that our qualitative data, while limited, offers some useful insight into items from the self-efficacy scale, which would be most imprecisely interpreted by students in our target population. Although some of the items retained in the scale could be also interpreted in several ways, at varying degrees of understanding; our interview data demonstrates that most students had a consistent grasp of what the task meant. Researchers concerned with the interpretability of this and other self-efficacy scales in chemistry could extend upon these findings by conducting interviews with different groups of students. Students with diverse chemistry backgrounds and experiences, as well as those whose first language is not English may provide perspectives on items that are not obvious to the developer and users of the scale. This depth of information should be regarded as absolutely vital for the validity of any inferences drawn from scores from the scale.

Implications and Future Research

Our interest when designing this study stemmed from our desire to study student motivation in chemistry. Due to the limited work in this area, our first step was to evaluate a set of modified items and revised scales, so that various aspects of motivation could be measured for a chemistry-specific population. Time constraints can often prevent instructors from administering lengthy scales or instruments, therefore, we tried to compile scales that would provide a balance between useful data and classroom administration time. Furthermore, student participation and completion rates tend to be lower with longer, more time-

consuming instruments (Lichtenstein *et al.*, 2008; Heredia & Lewis, 2012). Additionally, it is crucial that each scale is actually measuring what it's developers have purported it to be measuring. Therefore, it is incumbent upon researchers to thoroughly examine relevant psychometric evidence of such scales prior to, or as part of, their use. With the present work, we show that each of our revised scales to measure self-efficacy, interest, and effort beliefs demonstrate acceptable psychometric properties for use in a general chemistry setting. Additionally, our revised scales, which measure three well-defined latent traits, are comprised of a small number of items. Our revised scales consist of 7 initial interest, 8 maintained interest, 6 effort beliefs, and 6 self-efficacy items. When used together in future studies, this equates to 19 (with initial interest) or 20 (with maintained interest) total items to address three latent traits. By comparison, the CAEQ (measuring three distinct latent traits), and the CSCI (measuring five types of self-concept) are comprised of 69 and 40 items, respectively (Dalgety *et al.*, 2003; Bauer, 2005).

Our next step is to now utilize these scales to study the impact of teaching practice on students' self-efficacy, interest, and effort beliefs. Many studies are suggestive of the powerful influence of self-efficacy on performance (Zimmerman *et al.*, 1992; Pajares, 1996; Zusho *et al.*, 2003). To further corroborate these findings, and expand upon them in the college chemistry setting, studies involving measures of self-efficacy and performance together are needed. Even less explored are the relationships among interest, effort beliefs, and performance. There is a particular lack of research involving these latent traits in

college level sciences. In future studies, we plan to investigate how different practices affect these motivational factors as well as how these factors affect each other and ultimately students' performance. Therefore, follow-up studies will use *a priori* path analysis models to evaluate the correlation between scales and their mediation of course performance (Xu *et al.*, 2013).

In addition to our ongoing studies, we offer a few avenues for extension of the current study. A large-scale study could further utilize the power of structural equation modeling through invariance (or measurement equivalence) analysis. Invariance analysis allows the researcher to test whether a proposed model is equivalent across different groups of participants. One form of testing for invariance is to use a multi-group CFA, whereby the researcher is able to test for the equivalence of the measurement and structural solution (Brown, 2006). Put simply, invariance analysis can inform the researcher as to whether the same trait is being measured across different groups (race, gender, major), which is an important consideration for the validity of an instrument (Hutchinson *et al.*, 2008).

Instructors who are interested in gauging the motivational atmosphere of their classes might find our chemistry-specific scales useful. Students' beliefs about motivation and effort precede and govern their actions in the course. At the start of a semester, an instructor might want to have knowledge of his or her students' interest and confidence toward chemistry for the purpose of tailoring certain aspects of the course to their group of students. Due to the brevity of the scales, data could be collected at multiple time points throughout the semester with minimal time commitment. This could be especially useful if an innovative

instructional strategy were to be implemented. The instructor could evaluate the impact of their instructional strategy on dimensions beyond course performance measures. This would be informative as performance measures alone tell instructors nothing about a student's motivational or affective disposition toward the course, which are vital components of a student's academic success (Zusho *et al.*, 2003). We encourage educators who employ novel instructional strategies to consider measuring the motivational and affective processes of their students, in order to add to the current understanding of the impacts of these strategies. As stated in the 2012 DBER report, "the interplay between faculty behavior [i.e., teaching strategies] and student affect merits further exploration." We feel that the present work aids instructors in this exploration by providing measurement tools adapted for college chemistry, founded on prevailing theories from educational psychology, and subjected to the rigor of a thorough psychometric evaluation.

CHAPTER V

ARTICLE 2: CONNECTING ACHIEVEMENT MOTIVATION TO PERFORMANCE IN GENERAL CHEMISTRY

(This manuscript will be submitted for publication with co-authors Michael Phillips and Jack Barbera)

Abstract

Student success in chemistry is inherently tied to affective and motivational processes. We investigated three distinct constructs tied to motivation: self-efficacy, interest, and effort beliefs. These variables were measured twice over the course of a semester in three sections of a first-semester general chemistry course ($n = 143$). We explored the connections that exist among these three constructs as well as their connections to course performance. Multiple regression and path analysis revealed that self-efficacy measured during week 12 was the strongest predictor of final course grade followed by maintained situational interest. We also report that initial personal interest is a significant predictor of future self-efficacy. Our results have important implications by identifying variables related to motivation that have a significant connection to course performance among chemistry students. We briefly address how these variables could be targeted in the classroom.

Introduction

Achievement motivation is a multi-faceted and complex nexus of interconnected processes. In the context of education, the importance of motivation and affective processes cannot be overstated. Over the last 50 years in the field of psychology, much work has been dedicated to the understanding of these processes in terms of what drives students' choices and persistence in education (Weiner, 1990a). Many theories have been postulated over the years and with them, specific psychological constructs have been defined and operationalized in a myriad of studies. In chemistry education, research has centered on attitudes (Barbera, Adams, Wieman, & Perkins, 2008; Bauer, 2008; Xu & Lewis, 2011) and several motivational beliefs and processes including self-efficacy (Dalgety, Coll, & Jones, 2003; Smist, 1993; Villafane, Garcia, & Lewis, 2014; Zusho, Pintrich, & Coppola, 2003) interest (Dalgety & Coll, 2006b; Nieswandt, 2007; Uzuntiryaki & Aydin, 2009), self-regulation (Black & Deci, 2000), and self-concept (Bauer, 2005; Lewis, Shaw, Heitz, & Webster, 2009).

To better understand academic motivation as a whole, constructs should be measured and studied together with the intent of establishing connections between them (Bathgate, Schunn, & Correnti, 2014). In addition, knowledge of how they evolve over time is valuable for instructors interested in improving the motivational climate of their classrooms. In a prior study, we evaluated the psychometric properties of four scales that measured self-efficacy, initial interest, maintained interest, and effort beliefs (Ferrell & Barbera, 2015). In the present study, we extend upon our previous work by investigating the connections

between these constructs and how they predict course performance in a first-semester general chemistry course.

Background

Connections among motivational constructs. Self-efficacy, interest, and effort beliefs represent three distinct psychological constructs that can have an effect on motivation. In our previous publication, we provide a detailed description and literature review on each construct (Ferrell & Barbera, 2015). To date, no studies in secondary or post-secondary education have been found that link all three constructs together, either empirically, or theoretically. This is not surprising as effort beliefs research, in particular, is relatively sparse in the literature. However, several empirical studies have included two of the three constructs as measured variables.

Lent, Brown, and Hackett (1994) applied social cognitive theory to career development by formulating a model that included self-efficacy, academic interest, choice, and performance. Central to their model was the notion that self-efficacy is a major mediator of choice and development, and guides one's decision-making. The authors reviewed 13 relevant studies and found that of all the correlations with self-efficacy, interest was the strongest ($r = 0.53$), and performance was moderately correlated ($r = 0.38$). More recent studies have also corroborated these findings (Larson, Stephen, Bonitz, & Wu, 2014; Lee, Lee, & Bong, 2014; Lent et al., 2001; Lent et al., 2008; Smith & Fouad, 1999). The temporal ordering of self-efficacy and interest has proven more difficult to deconstruct. Lent et al. (2001), along with others (Lent et al., 2008; Silvia, 2003),

have found evidence that self-efficacy is a causal precursor to career-related interest. Thus, a students' confidence in a particular domain may lead to the development of interest. Others have suggested that self-efficacy and interest are reciprocally related (Nauta, Kahn, Angell, & Cantarelli, 2002), meaning that self-efficacy leads to interest just as much as interest leads to self-efficacy. Still others have produced viable models of interest and self-efficacy that are temporally equivalent with both being caused by variables not included in the model (Lee et al., 2014). In any case, these studies highlight the salience of considering interest and self-efficacy in achievement motivation models, as well as career and college major choice models.

The empirical research on effort beliefs remains sparse, and as a result, there is little evidence that addresses connections with other motivational constructs. Nevertheless, a few studies do exist that have measured self-efficacy or interest in combination with effort beliefs (Abdullah, 2008; Jones, Wilkins, Long, & Wang, 2012). Abdullah (2008) found a strong, positive correlation ($r = 0.51$) between positive effort beliefs and self-efficacy with a sample of students. This is consistent with Bandura's (1997) prevailing theoretical model, which suggests people with high self-efficacy tend to display more effort and persist longer with tasks than those with low self-efficacy. Although a display of effort is not equivalent to believing that effort will lead to positive outcomes, we argue that positive effort beliefs toward a task likely precede the exertion of effort. The relationship between interest and effort beliefs was investigated in one study on ninth grade math students (Jones et al., 2012).

Although the path models tested in the study included no significant causal relationship between interest and effort beliefs, data was presented demonstrating a moderate, positive correlation between effort beliefs and interest as well as a significant covariance among the residual terms.

Academic achievement and self-efficacy. Self-efficacy is the most widely studied of these constructs in terms of the link with academic achievement. Self-efficacy has been shown to consistently display a positive relationship with academic performance (Lightsey, 1999; Multon, Brown, & Lent, 1991; Robbins et al., 2004). Regardless of ability level, researchers have found that students who report high self-efficacy tend to outperform their peers who report low self-efficacy (Bouffard-Bouchard, Parent, & Larivee, 1991; Collins, 1982 as cited in Bandura, 1997). Students at the college level are no exception. The positive correlation between self-efficacy and academic performance has been observed repeatedly in college courses (Lent, Brown, & Larkin, 1986; Pajares & Kranzler, 1995; Pajares & Miller, 1994, 1995; Siegel, Galassi, & Ware, 1985; Zusho et al., 2003). Multon, Brown, & Lent (1991) conducted a meta-analysis on studies that measured college students' self-efficacy. They reported an average correlation of 0.38 between self-efficacy and academic performance, and that self-efficacy accounted for 14% and 12% of the observed variance in academic performance and persistence, respectively. A subsequent meta-analysis conducted by Robbins et al. (2004) surveyed 109 studies where various psychosocial and study skills variables were compared with academic performance. The two strongest psychosocial predictors of college GPA were

self-efficacy and achievement motivation (p s = 0.496 and 0.303, respectively). Furthermore, due to the non-compulsory nature of college, many studies investigating self-efficacy among college-age students have examined the predictive power of self-efficacy on persistence (Hull-Blanks et al., 2005; Lent et al., 1984; Vuong, Brown-Welty, & Tracz, 2010). Wright, Jenkins-Guarnieri, & Murdock (2012) sampled 401 first-year undergraduates and found that course self-efficacy measured at the end of the semester was a significant predictor of persistence to enroll in the second semester. This effect was found after controlling for relevant variables such as gender, high school GPA, and ethnicity.

Academic achievement and interest. Although there has been less research linking interest to academic achievement, a few notable studies should be mentioned. It is important to highlight that interest can be conceptualized as both a trait (personal interest) and a state (situational interest) (Hidi & Renninger, 2006). The following studies tested the connection between situational interest and course performance. Harackiewicz et al. (2000) investigated the short and long-term effects of college students' goal orientations, interest, and performance in an introductory psychology class. The authors utilized path modeling to test the causal ordering of the tested variables. Their results showed that in the short-term (one semester), students who reported higher maintained situational interest received higher grades in the course. Although this effect was significant, the authors point out that students had already received feedback from two exams prior to the measurement of their interest. Hence, the notion that level of interest accounted for the performance level could be muddled by the timing of

the measurements. In a related study, Hulleman et al. (2008) found that the utility value component of interest (i.e., importance for future), but not the intrinsic component (i.e., enjoyment of class), significantly predicted academic performance in college psychology. In this study, the interest was measured prior to any exams, eliminating the problem of timing for causal ordering.

Academic achievement and effort beliefs. Very little research has been published with effort beliefs as a measured variable. Among those studies that do exist, no direct link between effort beliefs and academic performance has been established. Instead, Blackwell, Trzesniewski, & Dweck (2007), in a study on sixth and seventh graders, identified positive effort beliefs as an important mediator between incremental theory of intelligence (Dweck, 2012) and positive learning strategies, which predicted grades. In a subsequent study, Jones et al. (2012) sought to replicate the model proposed by Blackwell et al. (2007) with ninth grade math students. They also found that positive effort beliefs mediated the relationship between incremental theory of intelligence and positive learning strategies, which were predictive of current grades. Although limited in scope, the findings of these studies expand upon the importance of considering effort beliefs as a mediator between implicit theories and adaptive learning patterns that lead to higher achievement (Tempelaar, Rienties, Giesbers, & Gijsselaers, 2015).

Factors linked to achievement in chemistry. Prior studies in chemistry have demonstrated that cognitive variables such as spatial skills (Carter, LaRussa, & Bodner, 1987), math ability (Lewis & Lewis, 2007; Spencer, 1996), and prior conceptual knowledge (Xu, Villafane, & Lewis, 2013) are linked to

achievement. For example, Lewis and Lewis (2007) were interested in predicting at-risk students in general chemistry. They found that a significant amount of the variance in students' ACS exam scores was accounted for by their SAT math scores. This suggests that students' math ability plays an important role in their success in general chemistry. In a related study, House (1996) identified that students' beliefs about their math ability was the strongest predictor of achievement in a college introductory chemistry course, even when ACT composite scores were considered. While cognitive abilities are an integral part of academic performance in chemistry, one cannot ignore the underlying non-cognitive beliefs and processes. Several non-cognitive factors such as self-efficacy (Zusho et al., 2003), attitude (Xu et al., 2013), self-concept (Lewis et al., 2009), and utility value (Gonzalez & Paoloni, 2015) have been correlated with achievement in chemistry courses. For example, Zusho et al. (2003), in a study with college chemistry students, found that self-efficacy accounted for the most variance in course performance, even after controlling for prior achievement. These studies underscore the fact that non-cognitive beliefs and processes are vital to student success in chemistry.

Present Study

Our aim in this study was to explore the possible relationships that exist between self-efficacy, effort beliefs, initial personal interest, and maintained situational interest as well as their connection to final grades in an introductory chemistry course. Toward this end, we used multiple regression and path analysis to test for plausible models that best represent these relationships.

Previously, we adapted and modified four scales intended to measure these motivation-related variables in a chemistry setting (Ferrell & Barbera, 2015). We demonstrated that all four scales produced sufficient evidence for validity and reliability with both an initial and cross-validation sample of first-semester general chemistry students, a necessary precursor to any further studies (Arjoon, Xu, & Lewis, 2013; Barbera & VandenPlas, 2011; Brandriet, Ward, & Bretz, 2013). Our data collection and subsequent analyses for the present study were guided by the following two research questions:

- (1) What are the connections among self-efficacy, interest, and effort beliefs with general chemistry students?
- (2) To what extent do self-efficacy, interest, and/or effort beliefs predict course performance in general chemistry?

We hypothesized that self-efficacy and the value component of interest would be the best predictors of course performance in line with what others have reported (Bong, 2001; Hulleman, Durik, Schweigert, & Harackiewicz, 2008; Pajares & Miller, 1995). In addition, we hypothesized that initial personal interest would predict maintained situational interest, a relationship that has been theoretically established and tested in several studies (Harackiewicz, Durik, Barron, Linnenbrink-Garcia, & Tauer, 2008; Hidi & Renninger, 2006; Nieswandt, 2007). We believe the results from this study will provide valuable insight for instructors interested in measuring the motivational climate of their classrooms as well as those who want to expand the evidence of impacts related to curriculum changes.

Methods

Participants

Participants for this study were recruited from three sections of a first-semester general chemistry course at a mid-sized Rocky Mountain region university during the fall of 2014. This is a required course for several science and health-related majors, and is the first in a two-semester sequence of general chemistry.

Course overview

Two different instructors taught the three sections during this term. Instructor A taught two sections and instructor B taught one section. These instructors were not only selected based on their willingness to participate in the study, but also on their backgrounds and similarities in course structure and teaching style. Each instructor had over five years experience teaching within the general chemistry sequence and had been recognized for their excellence in teaching the course. The instructors worked together, meeting regularly over the semester, to coordinate on a number of aspects for the courses. In addition to using the same textbook and homework system, the instructors agreed on the timing of assessments and weight percentages of assignments. Each course had an equal number of online homework assignments, weekly quizzes, and hour exams. While the content coverage of each was similar, based on the pacing of each instructor, they did not use matched homework or assessment items. Both instructors administered an American Chemical Society examination for their final exam. Lecture was the main teaching style for all three sections. While both instructors supplemented lectures with brief activities, problem solving tasks, and

virtual or physical demonstrations, neither would categorize their style as active learning.

Measures

Initial and maintained interest. To assess interest at two time points during the semester, an initial interest and a maintained interest scale were needed. Both scales were originally developed by Harackiewicz et al. (2008) and very similar items were used in a subsequent study by Linnenbrink et al. (2010). We adapted the items from both scales for chemistry and found the psychometric properties of each to be acceptable (Ferrell & Barbera, 2015). The initial interest scale is comprised of seven items, three of which are feeling-related items, and the remaining four are value-related. The maintained interest scale is comprised of eight items, with an equal split of feeling-related and value-related items. The initial interest scale is designed to measure the personal interest that students already have prior to taking the course. The maintained interest scale, on the other hand, is comprised of items that assess students' situational interest, which is supported by the structure of the course (Linnenbrink-Garcia et. al., 2013) Both scales are measured using a five-point Likert scale ranging from "strongly disagree" to "strongly agree".

Self-Efficacy. To measure self-efficacy, select items from the College Chemistry Self-Efficacy Scale were used (CCSS; Uzuntiryaki & Aydin, 2009), all of which were previously tested for adequate psychometric properties. We originally tested a scale that consisted of eight items, but two items were found to produce negative impacts on the models we tested, and were thus removed

(Ferrell & Barbera, 2015). The revised scale used in the present study consists of six items that ask students to rate how well they could complete tasks that would generally be encountered in an introductory chemistry course. The items are measured on a five-point Likert-type scale ranging from “very poorly” to “very well”.

Effort beliefs. To measure students’ effort beliefs about chemistry, we adapted a scale for chemistry originally developed by Dweck and Sorich (1997), subsequently published in Blackwell’s (2002) dissertation, and used in studies by Blackwell et al. (2007) and Jones et al. (2012). The original scale consisted of nine items, but our previous study produced evidence suggesting that the scale functioned better if three of the items were removed (Ferrell & Barbera, 2015). Hence, in the present study, we used the six-item scale measured with a five-point Likert scale ranging from “strongly disagree” to “strongly agree”.

Academic performance. Although there are many ways to measure academic performance in a college classroom, course grade is the most common. We chose to use course grade percentage because it represents the final outcome of the course and is the highest stake for students. Students with incomplete grades were not included in the analyses.

Data Collection

Data for this study were collected at two time points during the lecture period of each section. Time 1 (T1) data were collected during the first week of the semester and time 2 (T2) data were collected during the 12th week of the semester. On the days when data were collected, the researcher made an

announcement informing the students that their participation was voluntary and that their identities would be kept completely confidential. The students were then provided the survey packet along with a consent form that was approved by the Institutional Review Board. Students were given approximately 10 minutes to complete all of the scales (19 items at T1 and 20 items at T2) and seven demographic items (only given at T1). Due to the nature of this study, only students who completed all items at both time points were included in subsequent analyses.

Data Analysis

Descriptives. Descriptive statistics including mean, standard deviation, skew and kurtosis values were determined for all scales at both time points. We considered acceptable skew and kurtosis values to be -1 to +1 (Huck, 2012). Statistical Package for the Social Sciences (SPSS) 20.0 software was used for these analyses.

Multiple regression. Multiple regression analysis is a method used to predict or explain the variation in a dependent variable by examining its relations to several independent variables (Pedhazur, 1997). By utilizing multiple regression analyses, the researcher is able to identify and separate the individual effects of distinct independent variables on a dependent variable. Multiple regression analysis was employed in this study to examine to what extent self-efficacy, interest, and effort beliefs predict final grades. All necessary assumptions were examined including univariate normality, linearity, homoscedasticity, and normal distribution of residuals. In addition,

multicollinearity was checked among all of the independent variables. Finally, Mahalanobis distances were used to screen for potential outliers and influential points.

Path analysis. Path analysis is similar to multiple regression, but allows the researcher to estimate the effects of independent variables on each other as well as on the dependent variable and the causal directions associated with these effects (Pedhazur, 1997). A related technique, structural equation modeling is often used to specify a full model with latent and observed variables along with a path diagram representing the interconnections between the variables (Gonzalez & Paoloni, 2015; Xu et al., 2013). However, path analysis has no latent variables and generates a path diagram with path coefficients based solely on observed variables with only one indicator (Raykov & Marcoulides, 2000). The observed variables that only have outgoing paths are referred to as exogenous variables, with the variance in these being explained by factors not included in the model. In addition, they are assumed to be measured without error. Although measurement error cannot ever be fully eliminated, the measures used here have been previously well supported with evidence of validity and reliability (Ferrell & Barbera, 2015). The paths which have at least one incoming path are referred to as endogenous variables. The variance of endogenous variables is assumed to be explained by only the other variables included in the model, plus a disturbance term, which is analogous to an error term (Streiner, 2005). One of the advantages of using path analysis versus multiple regression is the ability to test how well competing models fit a given data set. To do this, each model is

examined for overall fit. There are a myriad of fit indices that researchers use to support the fit of a model, but we will use only the most common for path analysis. Global model fit was tested using the χ^2 test for goodness-of-fit, in which the desired result is to retain the null hypothesis (Streiner, 2006). While this is a good indicator of how well a model fits the data, it can be influenced by large sample sizes, which often leads the researcher to reject the null hypothesis (Schermelleh-Engel, Moosbrugger, & Müller, 2003). Thus, other fit indices, which are less influenced by the size of the sample, are often used to further evaluate model fit. We chose to use three of these, Root Mean Squared Error of Approximation (RMSEA) (Steiger, 1990) Goodness-of-Fit Index (GFI) and the Comparative Fit Index (CFI) (Bentler, 1990). We describe the CFI and RMSEA in detail in separate publication (Ferrell & Barbera, 2015). Briefly, the acceptable values for RMSEA and CFI are < 0.08 and > 0.95 , respectively (Browne & Cudeck, 1992; Hu & Bentler, 1999). The GFI is an absolute fit index that compares the amount of variance and covariance explained by the hypothesized model to no model at all (Byrne, 2013). GFI values > 0.85 are considered indicative of acceptable fit, but > 0.90 indicates good fit (Schermelleh-Engel et al., 2003). To generate our path diagrams, a set of *a priori* path models were tested using LISREL 8.80 (Jöreskog & Sörbom, 2006) with the Maximum Likelihood (ML) estimator. An example path diagram (Model 1) is shown in Figure 2.

Results

Demographics

During the first week of classes, 299 participants (91% of enrolled students) turned in survey packets. Of these, 144 had complete data sets at the end of the semester, resulting in approximately 52% attrition from the study. The reasons for this were missing data at one time point or another (35%), withdrawal from the course (14%), or lack of consent (3%). Participants completed a demographic survey only at T1. Of the 144 participants included in the study, 73.4% were female, and 93.4% had taken high school chemistry. Most of students (78.3%) were freshman or sophomore level and 81.8% reported this class as being their first chemistry class in college. With regard to race/ethnicity, 73.4% were Caucasian, 18.2% were Hispanic, 3.5% were African American, 0.7% were Asian American, and 4.2% were categorized as Other. The breakdown of majors was as follows: 14% chemistry, 21% other science (biology, physics, mathematics), and 65% were non-science (nursing, sports and exercise science, other).

Data Screening and Descriptive Statistics

All data sets were screened for careless responses (i.e., selection of the same option for every item) and none were found. Descriptive statistics and internal consistency estimates (Cronbach's alpha) for the students' scores are reported in Table 8. Originally, 144 complete data sets were included, but one case was dropped (see below), resulting in a final sample size of 143. The skew and kurtosis values were within the range of -1 to +1, with the exception of the

initial interest value scale. As a result, the inferences drawn from the scores from the initial interest value scale should be interpreted with caution as multiple regression and maximum likelihood estimation are not robust to univariate nonnormality (Curran, West, & Finch, 1996; Osborne & Waters, 2002). The internal consistency analysis for each scale reveals that Cronbach's alpha (α) values are acceptable (> 0.70), with the exception of the effort beliefs scores at T1 ($\alpha = 0.68$). This low internal consistency was taken into consideration when interpreting the results below. The descriptive statistics reported here are comparable to those reported in our previous study using these scales, suggesting consistency in the measurement (Ferrell & Barbera, 2015).

Table 8. Descriptive statistics for each measured scale ($n = 143$)

Scale	Mean	SD	Skewness	Kurtosis	Cronbach's α
II – feel	3.17	0.92	-0.20	-0.19	0.91
II – value	3.79	1.06	-1.41	1.88	0.90
EB T1	4.04	0.53	-0.20	-0.64	0.68
SE T1	2.87	0.74	-0.38	0.23	0.88
MI – feel	3.24	0.98	-0.41	-0.24	0.94
MI – value	3.47	0.84	-0.38	0.08	0.85
EB T2	3.95	0.69	-0.71	0.24	0.83
SE T2	3.73	0.62	-0.68	0.80	0.86
Final %	76.65	12.40	-0.49	0.22	--

Note: II – Initial interest, EB – Effort beliefs, SE – Self-efficacy, MI – Maintained interest

Outlier screening revealed that one case was potentially influential due to the large Mahalanobis distance of 21.73. Mahalanobis distance follows a χ^2

distribution and values in excess of the critical value for a given number of predictors at a $p < 0.001$ are considered potential outliers (Fidell & Tabachnick, 2003). Our final models included five predictors at most, corresponding to a critical χ^2 value of 20.52. Hence, the case was removed from further analyses.

Multiple regression. Multiple regression analysis was conducted to determine the effects of initial interest, maintained interest, effort beliefs, and self-efficacy on final course grade. For these analyses, all three sections were combined to make a single data set, consisting of 143 sets of scores. Correlations between all mean scale scores and final course grade are reported in Table 9. Tests for the assumptions of homoscedasticity, linearity, and normal distribution of residuals were conducted and met. The level of multicollinearity was checked and found to be acceptable based on low variance inflation factors (VIFs) (Pedhazur, 1997).

Table 9. Pearson correlations of all scales and final course grade

Scale	II-feel	II-value	EB T1	SE T1	MI-feel	MI-value	EB T2	SE T2	Fpct
II-feel	-								
II-value	0.674	-							
EB T1	0.426	0.628	-						
SE T1	0.312	0.227	0.104*	-					
MI-feel	0.475	0.182	0.241	0.227	-				
MI-value	0.342	0.208	0.194	0.238	0.657	-			
EB T2	0.173	0.136*	0.438	0.108*	0.602	0.497	-		
SE T2	0.335	0.288	0.332	0.443	0.530	0.490	0.497	-	
Fpct	0.201	0.142	0.112*	0.229	0.388	0.222	0.319	0.481	-

Note: II – Initial interest, EB – Effort beliefs, SE – Self-efficacy, MI – Maintained interest, Fpct – Final course percent. *Not significant at $p < 0.05$. All other correlations are significant.

For the preliminary model, final course grade was regressed on the T1 variables using the enter method. The overall model was significant

($F(4,138) = 2.70, p = 0.033, R^2 = 0.073, R^2$ adjusted = 0.046). The only significant main effect in the model was T1 self-efficacy ($\beta = 0.186, p = 0.033$).

In the next model, final course grade was regressed on the T2 variables. The overall model was significant ($F(4,138) = 12.83, p < 0.001, R^2 = 0.271, R^2$ adjusted = 0.250). The two significant main effects were feeling-related maintained interest ($\beta = 0.254, p = 0.021$) and T2 self-efficacy ($\beta = 0.404, p < 0.001$).

For the last model, final course grade was regressed on self-efficacy at both time points, and both initial and maintained feeling-related interests. We used a hierarchical regression method to test the incremental effects of T2 variables over the T1 variables in the model. The first model was significant ($F(2,140) = 5.362, p = 0.006, R^2 = 0.071, R^2$ adjusted = 0.058) with only one significant main effect, self-efficacy T1 ($\beta = 0.184, p = 0.034$). The second model was significant as well ($F(4,138) = 11.916, p < 0.001, R^2 = 0.257, R^2$ adjusted = 0.235, R^2 change = 0.186, $p < 0.001$). Both maintained interest feel ($\beta = 0.195, p = 0.039$) and self-efficacy T2 ($\beta = 0.374, p < 0.001$) were significant main effects in the second model. Neither of the T1 variables accounted for a significant amount of variance in final course grade when the T2 variables were introduced into the model. This data suggests that variables measured later in the semester account for significantly more variance in final course grade than T1 variables, even after the variance from T1 variables was accounted for.

Path analyses. Path analysis was used to test a set of path models, which describe a network of relationships simultaneously. The possible

connections between self-efficacy, effort beliefs, and interest were explored along with their effects on students' final grades. As with the multiple regression, all models were initially tested with the entire data set ($n = 143$).

Table 10. Fit statistics for each model tested, bold indicates value within acceptable range for the type of statistic

Model	$\chi^2(df)$	p -value	RMSEA	CFI	GFI	R^2 *
1	121.7 (12)	< 0.001	0.26	0.66	0.80	0.27
2	122.5 (13)	< 0.001	0.25	0.66	0.80	0.29
3	21.6 (10)	0.02	0.09	0.96	0.96	0.33
4	21.2 (9)	0.01	0.10	0.96	0.96	0.33
5	13.2 (9)	0.15	0.06	0.99	0.97	0.34
5a**	28.9 (22)	0.15	0.07	0.98	0.95	0.24, 0.34
5b**	44.7 (31)	0.05	0.08	0.96	0.92	-
$\Delta\chi^2$	15.8 (9)	> 0.05	-	-	-	-

* R^2 reported as variance accounted for in final grade

**Models 5a and 5b represent the multiple groups analysis. In model 5a, all parameters were estimated for courses taught by both instructors. In model 5b, parameters from instructor B courses were constrained to those of instructor A. The fit statistics are the same regardless of which course is constrained. The R^2 values separated by a comma represent the variance accounted for in final grade for each instructor's course(s).

The first model (Figure 2) we tested included all T1 variables (excluding value interest) predicting their respective T2 variables (e.g., initial interest predicting maintained interest), and all T2 variables predicting final course grade. Value interest was excluded in the path analysis here due to its lack of salience in preliminary models and to increase parsimony in the final models. Model 1 demonstrates very poor fit to the data, as shown in Table 10. Although the model predicts a significant amount of variance in final course grade ($R^2 = 0.27$,

$p < 0.01$), the χ^2 test was significant ($p < 0.001$) and both the RMSEA (0.26) and CFI (0.66) were out of their acceptable ranges. As with the multiple regression results, effort beliefs T2 was not a significant predictor of final course grade, but self-efficacy T2 and maintained interest feel were significant predictors ($p < 0.05$). We tested whether model fit improved when the path from effort beliefs T2 to course grade was removed (model 2), but no significant improvements were found.

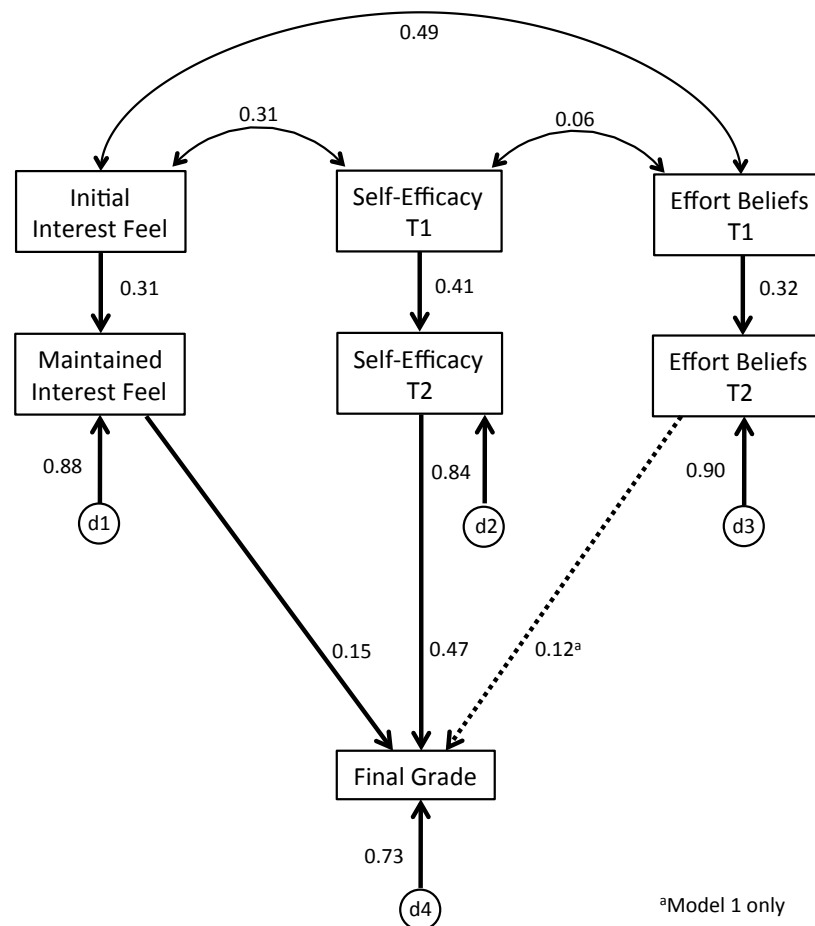


Figure 2. Path diagram of models 1 and 2. Standardized path coefficients are reported. All values > 0.14 are significant at $p < 0.05$. Dashed path from effort beliefs T2 to final grade was removed in model 2.

To explore the possibility of model misspecification, we examined the modification indices (MIs). Modification indices are equivalent to a one degree of freedom χ^2 statistic that can only be applied to parameters which are fixed. The MIs provide an estimate of how much the overall χ^2 statistic will drop when the parameter in question is freely estimated (Byrne, 2013). The MIs for model 2 indicated that correlating the disturbance terms of all three T2 variables would improve the overall fit of the model. By allowing the correlations of the disturbance terms to be freely estimated, we utilized a method known as correlated uniqueness (CU) (Kenny, 1979). Although somewhat contentious in the literature, CU allows the researcher to account for common method variance that can be particularly problematic when using the same survey and scale type to measure multiple variables (Marsh & Bailey, 1991; Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). When we re-specified the model to include these changes, the fit improved dramatically. While model 3 (Figure 3) still did not meet all of the necessary criteria for acceptable fit, the RMSEA value fell to 0.09, very close to the cutoff value of 0.08.

In addition to the links between pre/post scores of the same construct, we were interested in testing the predictive relationships between self-efficacy and interest. Toward this end, we tested competing models – the first of which had initial interest feel predicting self-efficacy T2 (model 4) and the second had self-efficacy T1 predicting maintained interest feel (model 5). Models 4 and 5 are identical to model 3 (Figure 3) with the exception of their respective added paths between interest and self-efficacy. Both of these relationships are supported by

the literature, in addition to a reciprocal relationship between self-efficacy and interest, as explained above. However, in the context of chemistry, we are unaware of any evidence that supports either hypothesis, therefore, both paths need to be tested.

Model 4 (Figure 3) added little improvement to overall fit from model 3, as evidenced by the significant χ^2 statistic ($p = 0.012$) and the RMSEA value (0.10). Furthermore, the path coefficient ($\beta = 0.04$) between self-efficacy T1 and maintained interest feel was not significant at $p < 0.05$. This suggests that self-efficacy scores at the start of the semester do not predict future interest in chemistry. All other significant paths from model 3 were similar in direction and magnitude in model 4.

When a path was added from initial interest feel to self-efficacy T2 in model 5 (Figure 3), the fit improved significantly. The χ^2 statistic was not significant ($p = 0.154$) and the RMSEA value dropped to an acceptable level (0.06). In addition, the path coefficient ($\beta = 0.22$) between initial interest feel and self-efficacy T2 was significant ($p < 0.05$). This finding provides evidence that interest in chemistry at the start of a course is an important factor in predicting students' self-efficacy at the end of the course.

We acknowledge that there could be a causal link between self-efficacy and effort beliefs. A prior study by Abdullah (2008) indicated a significant correlation between these two variables. We chose not to investigate this relationship because we did not have any prior theoretical backing to support the directionality of the relationship. Additionally, we did not feel that our sample size

was large enough to make any claims of causality without any theory to support them.

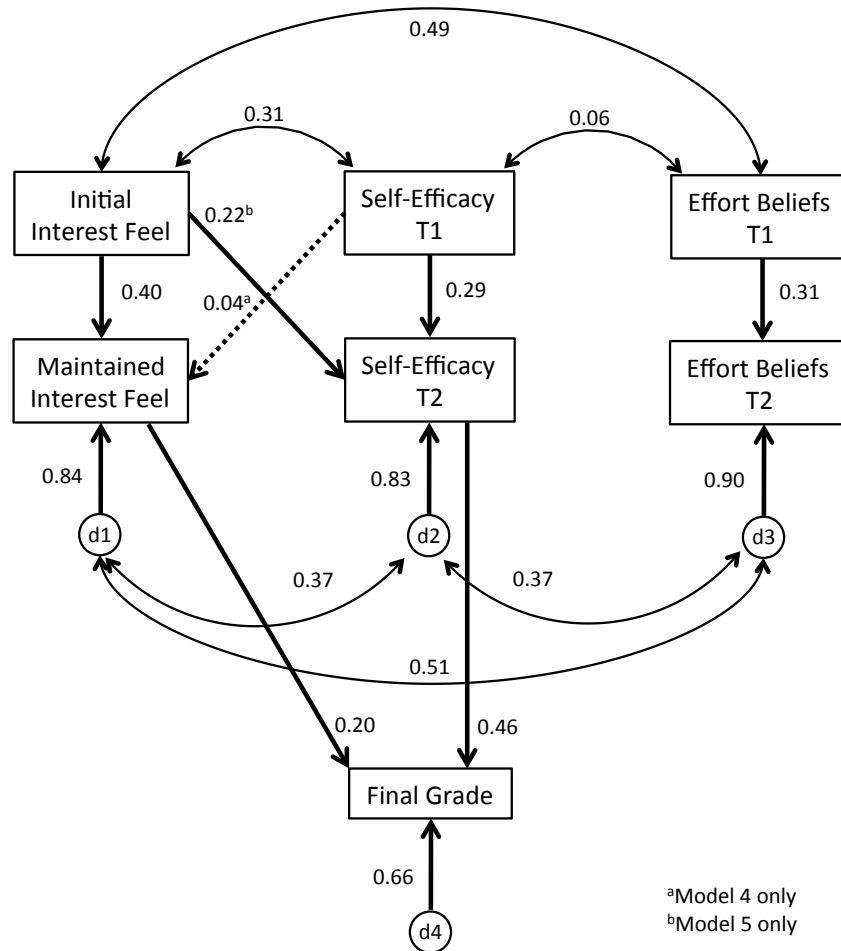


Figure 3. Path diagram of models 3-5. Standardized path coefficients are reported. All values > 0.14 are significant at $p < 0.05$. The paths between interest and self-efficacy are additions in models 4 and 5.

To investigate the possibility that the way the instructors taught influenced students' scores and the subsequent connections among the different variables, we performed multiple groups path analysis on model 5. Multiple groups analysis, analogous to testing measurement invariance (Cheung & Rensvold, 2002; Xu, Kim, & Lewis, 2016), is a method to test whether the grouping of

scores (e.g., by class, gender, race) generates significantly different parameter estimates and fit in a given model. In our case, the two sections taught by instructor A were pooled and had a sample size of 92. The section taught by instructor B had a sample size of 51. Multiple groups analysis utilizes a χ^2 difference test ($\Delta\chi^2$) as a means of identifying group differences in overall fit (Cheung & Rensvold, 2002; Scientific Software Incorporated, n.d.). Briefly, the model parameters were estimated for the two groups separately which were incorporated into a single χ^2 goodness-of-fit statistic (model 5a). Next, the parameters of the second group were constrained to those of the first and a new χ^2 goodness-of-fit statistic is generated (model 5b). The difference between the two χ^2 values is compared to a critical value using the difference in the degrees of freedom between the original and constrained models. We used an alpha value of 0.05 to assess significance of the differences between the two models. The difference in the χ^2 values was 15.8 and was not significant when compared to a critical value of 16.9. Hence, this test provides evidence to support grouping together the three sections into one data set (in model 5), due to similarities in the patterns of scores among students in all sections.

Discussion

In this study, we explored the connections between self-efficacy, initial interest, maintained interest, and effort beliefs as well as their impact on course grade in a first-semester general chemistry course. Participants were evaluated on these measures at two time points (week 1 and week 12). This work builds on our prior study (Ferrell & Barbera, 2015) by investigating the temporal links and

interconnections, via path analysis, that exist between these three motivational variables. In addition, we analyzed the extent to which students' grades were predicted by the path models tested. In our best-fitting model (model 5, Figure 3), we found that T1 variables were good predictors of their T2 counterparts and that initial interest (feeling) was a good predictor of self-efficacy T2. Additionally, 34% of the total variance in course grade was accounted for by the variables included in model 5.

Multiple Regression Analysis

To begin, we regressed final course grade on all measured variables at both time points. Based on prior research, we expected self-efficacy and interest to be potential predictors of course grade (Gore, 2006; Harackiewicz et al., 2000; Smist, 1993; Zusho et al., 2003). Our results are concurrent with this, suggesting that self-efficacy, measured at the end of the semester, is the strongest predictor of course grade, followed by feeling-related maintained interest, which also accounts for a significant portion of the variance in course grade. These results were also obtained when initial interest feel and T1 self-efficacy were accounted for in a hierarchical regression model. Effort beliefs was not found to be a significant predictor of course grade in any of our regression models. This is in line with what has been previously reported (Blackwell et al., 2007; Jones et al., 2012). However, effort beliefs T2 did exhibit strong correlations with all other T2 variables (see Table 9), demonstrating a clear relationship with distinct, motivation-related constructs. To further understand effort beliefs as a

psychological construct, we are conducting a qualitative investigation among general chemistry students that will be reported at a later date.

It is important to note here that the timing of measurement should be a central consideration when assessing students' motivation. At the start of the semester, self-efficacy T1 was a very weak predictor of final course grade and initial interest was not a significant predictor. By the end of the semester, self-efficacy T2 and maintained interest feel accounted for a much larger portion of the variance than the T1 variables. The predictive effect of self-efficacy measured in the middle of the semester, versus the beginning, on performance has been observed consistently by others (Bong, 2001; Gore, 2006; Lee et al., 2014). We postulate this could be due to two reasons. First, students are not well calibrated in their confidence when they walk in the door on the first day of chemistry. Hence, they may under or over-estimate their capability to complete certain tasks they will encounter in the course. Others have reported evidence of high school students' overconfidence in math, but suggested that college students would likely have better calibration (Pajares & Kranzler, 1995). Although most of the students in this study had chemistry in high school, some did not, which would further complicate the issue of calibration because of the lack of knowledge, skill, and prior attainments (Gist & Mitchell, 1992; Pajares, 1996). By the end of the semester, students had more recent experience with the tasks included in the self-efficacy scale and were better able to self-appraise their abilities (Wright et al., 2012). Bandura (1986) also notes that self-appraisals of ability should improve with time. Secondly, students, by the end of the semester,

had considerably more performance feedback to draw upon and with which to match their appraisals. Thus, it is likely that their confidence would better line up with their actual performance thereby improving the correlation between self-efficacy and final course grade. To begin accounting for this, we suggest that future classroom studies consider collecting a set of data after students have had some graded performance feedback. This can then be compared to data collected at the end of a course to get a better gauge of the impacts of the classroom environment. Additionally, we encourage future qualitative studies of the reported reasons for students' change on these variables and which aspects of the environment (if any) are reflected on as the nature of the change.

Path Analysis

We were not only interested in total variance explained, but also the predictive power of the independent variables in our model. Our first model (Figure 2) included all T2 variables as endogenous variables predicting final course grade and all T1 variables as exogenous variables predicting their respective T2 variable. This path model resulted in poor fit and fit improved little when the non-significant path from effort beliefs T2 to final course grade was removed (model 2). It was only when we allowed the disturbance terms of each T2 variable to be correlated that the model fit improved significantly (model 3). Correlated disturbances (or residuals) represent shared variance between two variables included in a model and some cause outside the model (Landis, Edwards, & Cortina, 2009). Often modification indices will be generated that estimate improved model fit should certain residuals be allowed to correlate. It

should be noted, however, that modification indices are not a substitute for substantive theoretical backing. Rather, they can be viewed as suggestions that should be interpreted in light of prior research and accepted theory.

Though there is little research on effort beliefs, Jones et al. (2012) found that effort beliefs and interest shared a significant amount of residual covariance in their model. In their model, both interest and effort beliefs had a common predictor – incremental theory of intelligence. Our model did not include incremental theory of intelligence (see Dweck, 2012), which could account for some of the shared residual covariance between interest, effort beliefs, and self-efficacy. Dweck (2000) argues that confidence (related to self-efficacy), while a valuable asset and predictor of achievement when things are going well, is not sufficient to carry students through difficult transition periods during their academic years (e.g., transition to junior high, or college). Rather, students endorsing an incremental theory of intelligence (related to positive effort beliefs) are more likely to persist in difficulty, whether they have high or low confidence in their current ability or intelligence. On the other hand, those with an entity theory and low confidence are more likely to lose ground when faced with obstacles by blaming their fixed intelligence rather than effort (Hong, Chiu, & Dweck, 1995). Although not a direct connection between effort beliefs and self-efficacy, this argument presents an alternative perspective of the grounding effect that an incremental theory of intelligence is thought to have on students with both high and low confidence. Thus, it is possible that self-efficacy, interest, and effort beliefs might have a common cause –implicit theories of intelligence. What a

student believes about the nature of intelligence and how intelligence grows influences the goals and behavior of that student (Dweck, 1996). Although this is a plausible common cause, a separate study should be conducted whereby implicit theories of intelligence are included as a measured variable.

Finally, initial interest was determined to be a significant predictor of self-efficacy at T2. We noted above the mixed results obtained from others (Lent et al., 2008; Nauta et al., 2002) on the causal link between self-efficacy and interest. We tested both causal directions – the first, self-efficacy predicting maintained interest feel, the second, initial interest feel predicting T2 self-efficacy. Our results revealed that initial interest was a better predictor of future self-efficacy than self-efficacy predicting future interest. Put simply, students who come into chemistry with a fascination or positive feelings toward chemistry will tend to leave the course with more confidence. However, as Nauta et al. (2002) found, the directionality could change with subsequent measurements of interest and self-efficacy. Thus, we do not suggest our results are in conflict with what has been reported previously, but rather, a snapshot of what is certainly a larger motivational landscape.

Implications for Research and Instruction

There are several implications of this study with respect to chemical education research and instruction. While this study did not include designed attempts to positively impact the measured variables, our study highlights the salience of considering affective and motivational processes in the classroom as influencing factors of course performance. An instructor cannot control what

beliefs students hold when they enter the classroom, but our results suggest that instructors could target interest and self-efficacy in their teaching strategies, which could impact course performance. Bandura (1986) argues that of the sources of self-efficacy, authentic mastery experiences are the most important. Obviously, all instructors hope their students achieve these experiences, but perhaps designing a curriculum or series of interventions around this idea would promote students' confidence, leading them to more adaptive motivational processes and ultimately, to success (Bandura, 1997; Zusho et al., 2003).

Self-efficacy is not the only target variable our model suggests addressing for enhanced course performance. While individual interest is not usually associated with performance in education, our results suggest that students with higher levels of feeling-related interest by the end of the semester performed better than those with lower levels of interest. Several interventions and teaching styles have been developed with the aim of impacting students' interest (Häussler & Hoffmann, 2002; Hulleman, Godes, Hendricks, & Harackiewicz, 2010; Hulleman & Harackiewicz, 2009). Hulleman et al. (2010) incorporated a brief intervention in a college psychology class whereby students in the treatment group were asked to write a short essay about the relevance of a topic being covered to a significant person in their lives. This short assignment resulted in a significant effect, both on increased interest in the course and on their final grades. Furthermore, the effect was most pronounced for students with low expectancies for success in the course, who are likely most at-risk for failure. Social-psychological interventions such as these should not be written off as

hand waving or magical. There are substantial effects that have been observed for very simple, but carefully planned interventions. Yeager and Walton (2011) have this to say about social-psychological interventions,

brief exercises that do not teach academic content but instead target students' thoughts, feelings, and beliefs in and about school - have had striking effects on educational achievement over months and years.

Limitations of Study

Our study has several limitations that should be made known. First, our sample was derived from one institution during one semester and did not include all sections of the course during that semester. Of the original sample, less than 50 percent of the participants provided all of the necessary criteria for inclusion in the final analyses. Finally, those who were included were primarily female and Caucasian. Thus, the generalizability of our results to other populations of general chemistry students is limited.

Second, we recognize that our models are imperfect and that all models are mis-specified to some degree (MacCallum, 2003). Although an assumption of path analysis and multiple regression is that the independent variables are measured without error, this can never be fully achieved. As noted earlier, our previous study (Ferrell & Barbera, 2015) supports the quality of the measures in terms of validity and reliability. We thoroughly investigated the psychometric properties of each scale used in the present study, a critical first step to any substantive investigation using self-report data. By having measures that have been properly designed and target the variables appropriately is critical to reducing the effects of systematic measurement error. Despite our efforts, our

reliability estimates are imperfect, therefore the statistical power and parameter estimates in our model are incorrectly estimated to a certain degree (Cole & Preacher, 2014). Furthermore, we acknowledge that without a full structural model, we cannot account for the measurement error of individual items. While many others have used path analysis in a similar manner as we have (Bong, 2001; Jones et al., 2012; Lent et al., 2008), full structural models are superior for testing theory.

The models reported here are merely an estimation of parameters and do not necessarily accurately describe the motivational phenomena at play. Therefore, we encourage others to cross-validate our models using a different sample to ensure that the pattern of relationships observed here is not a capitalization on chance (Hermida, 2015).

Despite these limitations, we believe the work presented here is important for the chemical education community by adding to the burgeoning base of research on affective and motivational processes of chemistry students. We have presented the performance impacts of several motivational variables and highlighted the connections that exist between them. More research is needed to further our understanding of the complexities that exist with respect to academic motivation in the hopes of improving our curricula to enhance learning among future students.

CHAPTER VI
CONCLUSIONS, IMPLICATIONS AND
FUTURE RESEARCH

Conclusions

The research described herein covers an area of education that cannot be ignored when considering the importance of student learning in the chemistry classroom. Most often, cognitive variables such as problem solving or content knowledge receive the most attention in chemical education literature. Less studied, but equally as important, are the affective and motivational states of students. The level to which students are motivated will determine what they do with the skills and knowledge they possess. For instructors and educators to gauge where their students stand in regards to motivation or other affective traits, they must have quality measurement tools. The purpose of this study was to adapt and modify existing scales to measure self-efficacy, interest, and effort beliefs among students enrolled in introductory chemistry. The project consisted of three phases. The first two phases dealt directly with the scales and their use in the classroom, while the last phase was directed toward investigating the understudied construct of effort beliefs from a qualitative angle. This chapter summarizes all three phases of the study by addressing the research questions posed in the first chapter of this dissertation.

Q1 What modifications are needed to produce brief, chemistry-specific scales of self-efficacy, interest, and effort beliefs?

This part of the study laid the foundation for the remainder of the project. Before the scales could be used with any hope of producing valid and reliable scores, they had to first be adapted to fit a chemistry context, and then investigated for modifications that would produce the best data for the target population. Toward this end, the wording of the items was changed to reflect a chemistry setting (see Appendix C for all items). Next, the newly adapted scales were given to a large sample of general chemistry students and a sub-sample of that group was interviewed about the scales. As a result of the qualitative and quantitative threads, the number of items for two (self-efficacy and effort beliefs) of the four scales were reduced and the other two (initial interest and maintained interest) scales used in the study were split into two factors (feeling and value).

Q2 What evidence supports the functioning of each of the modified scales?

To gather evidence on whether the scales were functional among students in general chemistry, both qualitative and quantitative techniques were employed and the data gathered was analyzed. This was done to ensure that the conclusions drawn from the quantitative data were supported by the qualitative data and vice versa. After initial modifications were made following the initial data collection, a cross-validation study was conducted to further test whether the modified scales produced valid and reliable data in additional populations.

The quantitative techniques to test the functionality of the scales included confirmatory factor analyses (CFA), ANOVAs, and internal-consistency (reliability) tests. Each scale was tested individually with CFA. The CFAs were used to test whether a particular grouping of items was statistically reasonable and provided information for items that do not belong in that grouping. The CFA results from initial data collection (see Tables 1 and 2) suggested that three items from the effort beliefs scale and two items from the self-efficacy scale be removed. In addition, the results suggested that the CFA model would better fit the data if the two interest scales were split into their corresponding subscales (feeling and value). This split has also been corroborated by others (Linnenbrink-Garcia et al., 2010). All of the modified scales with factor loadings can be found in Figures C1-C6. The internal consistency (Cronbach's alpha) for each scale was found to be reasonable to good for the type of applications for which these scales would be used. The ANOVA tests offered further evidence that the scales were measuring the intended variables by suggesting that chemistry majors had higher interest and self-efficacy than non-science majors.

In conjunction with the quantitative data, the qualitative results suggested that some of the same items be removed for either ambiguity in the meaning or poor readability. While there was not complete overlap in the results from both the qualitative and quantitative threads, all items that were exceedingly problematic from either a model fit or reading comprehension standpoint were removed.

Following the initial round of modifications, the scales were administered to a second large sample of general chemistry students at two institutions. These data were subjected to CFA to cross-validate the model obtained from the initial data collection. The results (see Table 3) from this cross-validation demonstrated that the model fit was adequate and comparable to the fit obtained in the initial data collection, suggesting that the scale modifications were appropriate and acceptable.

Q3 To what extent do students' self-efficacy, interest, and effort beliefs change across the first semester of general chemistry?

To answer this research question, a series of paired samples *t*-tests were conducted using each scale individually. Only students who completed the scales at both time points, beginning and end of the semester, were included in this analysis. The results are summarized in Tables 4 and 5. The interest scores could not be directly compared across the semester because the items differed between the initial and maintained interest scales. However, for effort beliefs and self-efficacy, the same trends were observed in both the initial and cross-validation samples. On average, students' scores for effort beliefs dropped across the semester and students' scores for self-efficacy rose across the semester. These results suggest two things that changed from the beginning of the semester to the end. First, students felt more confident in completing tasks encountered in general chemistry. Second, students were less likely to believe they could improve their chances of success in the course through effort.

To test whether the change in self-efficacy scores across the semester depended on major, mixed between-within ANOVA tests were conducted (see

Table 7). The results demonstrated that every grouping of major (chemistry, other science, and non-science) increased in self-efficacy with similar effect sizes. This suggests that overall increase in self-efficacy for the entire sample also held for each grouping of major.

Q4 To what extent are students' self-efficacy, interest, and effort beliefs affected by brief interventions targeting their values and implicit theories of intelligence?

One of the goals of this research study was to provide quality measurement tools for constructs related to motivation such that an instructor could introduce a change to his or her curriculum and measure the impact of that change from a motivation standpoint. Hence, the second phase of this project included introducing different, brief interventions in two sections of general chemistry (treatments) while maintaining a normal curriculum in a third section (control). The interventions used in this study were inspired by the work of Blackwell and colleagues (2012), Nussbaum and Dweck (2008), and Cohen et al. (2009). A summary of all interventions along with a schedule of the interventions can be found in Appendix D.

A MANCOVA test was conducted to determine if there were any differences among students' self-efficacy, interest, or effort beliefs scores between the different sections. The covariates used in the MANCOVA were the time 1 scores at the start of the semester. As a result, the students in different sections could be compared by how their scores differed at the end of the semester while accounting for their scores at the start of the semester. The results indicated that there was a significant difference on at least one of the

measures between the three course sections, while accounting for the time 1 scores, $F(8,270) = 2.82$, $p = 0.005$, Wilks' $\Lambda = 0.852$, partial $\eta^2 = 0.077$. Wilks' Λ is the MANCOVA test statistic and in this case, it is significant. Eta-squared (η^2) is a measure of effect size, or the magnitude of a given effect on the dependent variables (Fay & Boyd, 2010). In this case, the effect is which section the students belonged to and the dependent variables are their mean scores (e.g., self-efficacy, interest, effort beliefs). A value of 0.077 means that 7.7% of the variance in mean scores can be attributed to the course section.

Following the significant MANCOVA, a series of planned contrasts were analyzed. Planned contrasts are *a priori* tests to compare groups based on some variable. Planned contrasts were used here to test if there were differences among the three sections based on the mean scores on any of the measures. Planned contrasts are superior to *post-hoc* analysis because the Type I error risk is reduced (Abdi & Williams, 2010). The results (Tables C4 and C5) revealed that students in either of the treatment sections scored higher on effort beliefs, maintained interest (feeling), and self-efficacy compared to students in the control section. No differences in students' scores were found between the two treatment sections. Thus, it can be concluded that students who were involved in brief interventions targeting values and motivation during the semester scored higher, on average, than those who were not on three of the four scales given at time 2. Although this is in no way a causal inference, it does point to the potential for finding differences in students' motivation based on instructional interventions.

Q5 What are the connections among self-efficacy, interest, and effort beliefs with general chemistry students?

To address this research question, multiple regression and path analysis was used on the data collected during phase two of this study. Scores for self-efficacy, interest, and effort beliefs were collected at two time points during the semester. Based on prior research, it was expected that the time 1 measures would predict the time 2 measures of the same variable (e.g., self-efficacy time 1 predicting self-efficacy time 2) (Bong, 2001; Harackiewicz, Durik, Barron, Linnenbrink-Garcia, & Tauer, 2008). In addition, there was evidence suggesting a causal link between interest and self-efficacy (Lent et al., 2008; Nauta, Kahn, Angell, & Cantarelli, 2002; Silvia, 2003). As there is some debate regarding the temporal ordering of these two constructs in the literature, models in both directions were tested (see models 4 and 5 in Table 10).

Our results suggest that all of the time 1 measures significantly predicted time 2 measures of the same variable (see Figure 3). Multiple regression analysis indicated that the value component of interest was not a significant predictor in the models. Hence, it was removed from any further statistical tests. Of the matched variables (e.g., self-efficacy time 1 and self-efficacy time 2), initial and maintained interest showed the strongest predictive relationship, meaning that students who begin the semester with interest in chemistry tend to maintain that interest throughout the semester. Both measurements of effort beliefs and self-efficacy showed this relationship as well, but to a lesser degree.

When the relationship between self-efficacy and interest was tested, the feeling component of initial interest was found to be a better predictor of time 2

self-efficacy than self-efficacy time 1 measure was of the feeling component of maintained interest. Also, the model fit was much better when initial interest was the predictor (model 5) as opposed to self-efficacy time 1 (model 4). Put simply, this means that students entering general chemistry with an interest in chemistry, on average, will be more confident in their ability in the course than those who are not interested at the start. Although this finding is theoretically sound, the model should be cross-validated with a larger sample to ensure proper validity.

Q6 To what extent do self-efficacy, interest, and/or effort beliefs predict course performance in general chemistry?

As with Q5, multiple regression and path analysis were used to address this research question. The interest here was both the overall variance predicted by the model and the variance accounted for by each individual predictor. It was expected that self-efficacy would be the best individual predictor of final course grade, based on prior research (Lent, Brown, & Larkin, 1984; Multon, Brown, & Lent, 1991; Robbins et al., 2004; Zusho, Pintrich, & Coppola, 2003). Furthermore, self-efficacy time 2 was expected to predict performance better than time 1, which has been supported by others (Bong, 2001; Gore, 2006; Lee, Lee, & Bong, 2014).

As expected, self-efficacy time 2 emerged as the strongest predictor of course grade in both the multiple regression and path analyses when all scales were included in the model at both time points. When only self-efficacy measures were added to the model, time 2 accounted for 18% of the variance beyond time 1. Maintained interest (feeling) was also found to be a significant predictor of course grade. Together maintained interest (feeling) and self-efficacy time 2

accounted for 18.6% of the variance in course grade above that of initial interest (feeling) and self-efficacy time 1 in a hierarchical regression model. For the path analysis, the final model tested (model 5) accounted for 34% of the variance in course grade. These results highlight the importance of motivation-related variables in education by demonstrating their strong connection with course performance in general chemistry.

Q7 What are the sources and influences of effort beliefs toward chemistry among general chemistry students?

To answer this final research question, twenty-one students were interviewed. Participants were either currently or previously enrolled in first semester general chemistry at the time of their interview. Fourteen of the participants had recently completed the chemistry course in the prior semester and had two sets of effort beliefs scores (pre and post-semester). The remaining seven participants were interviewed during the semester in which they were enrolled in the course. Participants represented a variety of majors including chemistry, pre-health majors, biology secondary education, sports and exercise science, nursing, physics, and others. Most of the participants were freshman and had taken high school chemistry, though there were a few who indicated they had not.

Participants were interviewed in a quiet room, free of distractions, and asked about their scores on the effort belief items as well as what they thought were the influences of those beliefs. For those who reported a different pattern of responses from pre to post-semester, they were asked why they thought their scores changed. A few notable themes emerged related to what or to whom had

helped to shape the participants' effort beliefs as well as what may have contributed to a shift across the semester. Students primarily responded to the question, "Where do your effort beliefs come from?", in one of two ways: my family/upbringing or my personal experiences.

The first theme that emerged was the relationship of the students' effort beliefs to their family, most notably their parents. Of the twenty-one participants interviewed, six of them sourced family and upbringing as the primary source of their effort beliefs. For example, one participant stated, "I was raised on the belief: 'What you put in is what you're going to get out'. So, that's the way I feel." Another participant cited his mother's experience as having shaped his effort beliefs,

I learned more about effort and hard work from my mom. She's a single parent, so I learned how to – hard work tends to prevail from talent and all that, and she worked a lot. So, I assumed that to get what I want and to help provide for my family, I'm going to have to work my butt off. So, I tend to put in 110% percent in anything I do.

When asked about why she would not give up unless it was her final option, one participant told a story about her father and her upbringing,

My parents because they have had a hard life. Me – when I came here [from Africa], I was like, "Oh, I'll just be a kid." But, I had to grow up really fast because my parents had to go to work at night, so I would have to stay up and watch my siblings until a parent got home. My dad – I think he had gotten a degree in Africa, but for some reason it didn't transfer so he had to re-do school. Yeah, he went to school and went to work at the same time. Plus, my parents and all my siblings – there's a total of nine of us in the house. And so, it's like they were always persistent at doing hard work and they're like, "We're going to try to make a better life." So it's kind of me trying to give back to my parents. When they get old, I want to be able to take care of them and not just put them in a retirement home.

These stories and quotes are evidence that students entering a chemistry course come with a foundational belief about effort that began long before they came to college. Dweck and colleagues have argued that children's implicit theories of intelligence, closely related to effort beliefs, are at least partially shaped by their parents (Dweck, Chiu, & Hong, 1995). In particular, the feedback parents give their children has a strong impact on how the child will view the controllability of his or her intelligence. Some research has suggested that parents who praise their children's processes ("good job") as opposed to their traits ("you're smart") tend to promote more adaptive motivational frameworks in their children (Gunderson et al., 2013). Certainly, each student has a unique upbringing that contributes to his or her personal philosophies and beliefs toward education and the evidence presented here points to the importance of family involvement in the development of effort beliefs.

A second theme that emerged when participants were discussing sources and influences of their effort beliefs was personal experiences, particularly those in educational settings. Instead of pointing to family members or mentors, most students cited their own experiences with working hard as the primary source for why they held certain effort beliefs. One participant put it simply, "Because every time that I ever put in effort into something, it gives me a good outcome." Another student remarked similarly when asked about where her view on effort originated, "Probably just from experience. It's [chemistry] not one of the subjects that just comes to me. But, it is one of the subjects that I can – I have been able to pass in the past and understand the concepts in the class. I just have to actually try."

These quotes point to the notion of past experiences in education directing future behavior. Past experiences are foundational for formulating beliefs and expectancies, which then set the course for enacting certain behaviors. Behavior and personal factors (beliefs, emotions, etc...) represent two of the three vertices in Bandura's reciprocal determinism triangle (Bandura, 1986; Pajares, 1996). Both behavior and personal factors reciprocally influence each other – the personal factors along with environment (third vertex) tend to elicit certain behaviors that are then reflected upon by the individual, which sets in motion a recalibration of personal factors that will affect future behaviors. One participant laid this out nicely as part of his personal beliefs on effort,

I do believe that everything – there's some sort of innate ability that everyone's going to have for it. But, if you work, you can definitely improve it. There's this guy that's got this cycle – if you work hard, you get better. And, if you do better, it becomes more fun. If it's more fun, you're more willing to put in more work. And, it's just that endless loop. So, I definitely see that. So, if there's a kink in there, like, sometimes with 112 [second semester general chemistry], I don't end up getting it. So, it becomes less fun, so I want to put in less work. I know that with 111 [first semester general chemistry], I just kept getting better and better at it. So, I wanted to put more work in.

This student was able to see how his beliefs about effort and improvement plays into the cycle of motivation and outcomes through self-reflection and behavior adjustments.

Some of the participants noted connections between their effort beliefs and both family involvement and personal experiences.

One, in particular, described how immovable her effort beliefs were, despite her lack of success during her first attempt at general chemistry. She stated,

It's [my philosophy] not going to change – which is – I think if you work harder at anything, you'll get better at it. I don't think there's a specific subject where if you work as hard as you can, that you can't get any farther...In everything, you can move forward. That's never going to change no matter if a chemistry class – one chemistry class doesn't go well. Like, my chemistry class now is going well and that's probably because I still have the same beliefs.

She went on to say that her beliefs are a result of how she was raised and remarked, “My mom believes that you work your hardest. And, when you work your hardest, you get to the level you want to be at.” Later, she reflected on a time in her life when she had to relearn several grades of schooling due to an accident, “I worked my hardest. I had to go all the way back where everyone else was still moving forward. I had to go back to 5th grade in the 8th grade. That's a big jump and relearn all of that in a year.” This participant was very aware of her struggles and the history of how effort has gotten her to where she is today, both of which shaped her strong, positive beliefs about effort. Another student pointed out the need to have personal experiences to bolster her parents' influence on her effort beliefs. When asked about the split between influences from her personal life versus her parents' example, she remarked,

I'd say like a third from my parents and maybe two thirds from my own experience - because them just telling me that, that's not going to make me see it. But me acting it out was like, 'Oh, okay if I work hard enough, it will be okay.'

Not every participant could be interviewed about how the chemistry course itself influenced his or her effort beliefs due to the timing constraints in the study. However, of those that were asked about their experience in the course, a few

had some noteworthy things to say. Two students, in particular, remarked about how their effort beliefs changed for the worse as a result of their experience in the course. The first student, whose mean score dropped from 4.3 to 2.3 across the semester remarked,

I loved the teacher, she was a fantastic person – I was meeting with her once a week and I was meeting with a tutor from the tutoring center. And, I did not do very well on the first exam, but I started doing better and then I kind of dropped off again even after meeting, working with a tutor. Like meeting with the teacher wasn't helping me as much as I thought it was going to. And so, I stopped meeting with her. It went from me going to get help when I needed it to me like, "nobody's helping me, they're not helping me the way I need them to help me." And so, I was putting in more effort and it wasn't coming back out.

It is not clear from the quote how much effort the student was putting in or what type of effort was being applied. However, there was a clear disconnect between her expectations for success, based on her effort, and the reality of her performance. In addition, she was displeased with the type of help she was receiving from her instructor and the tutor and this could have set in motion a belief that her efforts to achieve success were futile. As a result, her self-reported effort beliefs scores dropped significantly.

The second student that reported a drop in effort beliefs across the semester was influenced by her peers and a tutor to believe that she was "not a chemistry person". Regarding this notion, she reflected on her thoughts about herself,

And so, I guess I was like, "I'm not doing well because I'm not good at it" instead of just, "it's something that's not really clicking in your head, it's not your fault". It's just something that gave me more difficulty – like some people really don't understand biology, some people really don't understand English and how to interpret scripts and stuff. I guess I was just told why I was struggling so much – it wasn't in my control.

This student began to endorse an entity theory of intelligence toward the subject of chemistry (Dweck, 2012). She saw her ability in chemistry as something that was innate and out of her control, instead of something that could grow with effort. This is evident in her effort beliefs mean score drop from 3.8 to 3.0.

Not all shifts in effort beliefs scores were negative. One student in particular reported an increase from 3.3 to 4.5 across the semester. When asked why this shift occurred, he pointed to how he was rewarded by his dedication,

I just enjoyed it more when I applied myself. I did a lot better last semester than I did in high school. It's because I put the effort forth – so that kind of gave me a better feeling about it. I think, for me, it made me realize that if I work harder at it and do more problems, that's something I can overcome or accomplish. It's actually rewarding to get through a problem and get the right answer. It's [a] cool feeling, you know.

This student highlighted that effort and attainments of success can be a perpetual cycle, whereby effort begets success and then success begets more effort. Bandura (1997) argues that experiences of success are the most influential source of self-efficacy, and that self-efficacy influences the amount of effort one is willing to expend on a task. Those students who see the fruit of their efforts will tend to expend more effort in the future and also believe that their effort is not wasted.

This qualitative exploration has shed light on the sources students perceive to be most salient in shaping their personal beliefs about effort. Every student will come into general chemistry with a different upbringing and different experiences. As a result, they will formulate different beliefs about the fruitfulness of effort in their chemistry class. However, the observations presented here

suggest that there are two general common sources that most students seem to identify as influencing their effort beliefs – personal experiences and family interactions. Just as students take inventory of their educational experiences in classes prior to taking general chemistry, so will they take inventory of their experiences with effortful actions in general chemistry. These experiences will be reflected upon and used as a source of future beliefs about effort. Hence, it is incumbent upon general chemistry instructors to be mindful of this in the design of activities, homework, and assessments. Those students who have the most positive reinforcement for their effort tend to be those who exhibit the most adaptive motivational patterns in education.

Implications of Study

The research described here combined both quantitative and supporting qualitative methodologies to begin to address an important research problem in the Discipline-Based Education Research (DBER) communities. Retention of students in STEM disciplines is an issue to which many researchers have and are continuing to devote themselves. The 2012 DBER report clearly outlines the need to consider the affective dimensions of students, including motivation, when designing curricula in the STEM disciplines (National Research Council, 2012). In order to make judgments about students' affective and motivational states, researchers must have quality measurement tools in place as a means to collect data. This research project aimed to extend existing research on the measurement of motivation in education and connect it to college level introductory chemistry. This work can serve as a model for adapting and

adopting these, and similar, measures for other chemistry courses and within the other STEM disciplines.

The primary goal of this project was to provide the chemical education community with four brief scales that have been designed and modified to measure three salient motivational variables (self-efficacy, interest, and effort beliefs) among students in introductory chemistry courses. The adapted scales underwent a significant psychometric evaluation, including cross-validation, to ensure that the data produced would be valid and reliable. Although no scale or survey can be considered a perfect measure of a given variable, the results reported here demonstrate that the motivational scales show acceptable to good psychometric properties. Furthermore, the scales combined comprise only 19 items (pre-semester) and 20 items (post-semester), allowing them to be administered in under 10 minutes. This is an advantage for instructors where time for non-content related activities is sparse at best.

For instructors interested in alternative teaching strategies and interventions, data from the scales described in this project will add a dimension to the body of evidence that exists for the impact of these methods. Whether it is cooperative learning, flipped classrooms, or simply a one-day intervention targeting interest, the chemical education community will benefit by knowing how students' motivation is affected by whatever curricular change is undertaken. For example, it is possible that flipped learning could have little to no effect on students' grades. With only that information, many instructors would probably write off flipped learning as pointless. However, if the instructors also took an

inventory of the motivational climate in the flipped learning classrooms, perhaps they would find that students' interest or self-efficacy in chemistry grew compared to traditional instruction. This additional information would be valuable for a potential user of the flipped classroom model in weighing the benefits and limitations of this practice.

In addition to providing several new scales to the chemical education community, the results of this study also suggest possible connections that exist between the motivational variables measured and overall course performance. This is a very important dimension to the study as course performance is a central concern for chemistry instructors and is the primary outcome variable for most courses. As a result, instructors are better informed as to the potential implications of designing a curriculum to improve students' self-efficacy, for example. However, this study was only able to examine a small sample of students from one institution. Thus, the results, while valuable, may not represent the population of introductory chemistry students as a whole.

Finally, the results of a series of brief interventions in two general chemistry sections are reported in this study. These educational interventions represent something new to the chemistry education community. No published studies were found to include these or similar interventions being used in college-level chemistry. The results of this study suggest that students who participated in the interventions had higher self-efficacy, effort beliefs, and maintained interest (feeling) at the end of the semester than students who did not. Again, the generalization of this portion of the study is limited due to the sample size and

other constraints. Nevertheless, the results show that a complete course overhaul may not be necessary to enhance students' motivation. Rather, simple and brief interventions can have a powerful effect on students from a motivational standpoint.

Future Research

The possibilities for future studies involving the motivation-related scales modified in this study are many. Perhaps the simplest application of this research is for instructors at various types of institutions to use the scales in their own classrooms to gauge and track the motivational climate. The results reported here indicate that, on average, effort beliefs drop and self-efficacy increases as the semester proceeds. This may or may not be true for general chemistry students at other institutions. Along that same vein, researchers could take another approach by looking at more diverse populations. At institutions where the ethnic diversity is high, researchers might be able to observe different trends in scores for different ethnic groups. From a structural equation modeling standpoint, invariance analysis would be beneficial to the chemical education community (see Chapter IV). It is possible that the structural models reported in this study might not hold up the same way with different populations of students. Perhaps the meaning of the items would be interpreted in a different manner, leading to a different set of items than the one described here. This would be very important information to have, as college campuses will continue to become increasingly diverse.

The primary impetus behind this entire study was to provide quality measurement tools to chemical educators in order to evaluate alternative teaching strategies from a student motivation angle. There are many published studies of innovative and novel approaches to teaching general chemistry, as outlined in Chapter II. Most of these were not, however, examined through the lens of student motivation. The data generated from the scales described in this study would be a valuable addition to the body of research on any of the aforementioned teaching strategies. It is important that the community of chemical educators not only strive for better course performance in introductory chemistry courses, but also that they are made aware of the potential impacts of various teaching methods on student motivation.

REFERENCES

- Abdi, H., & Williams, L. J. (2010). Contrast analysis. In N. J. Salkind (Ed.), *Encyclopedia of research design* (Vol. 1, pp. 243-251). Thousand Oaks, CA: Sage.
- Abdullah, M. (2008). Children's implicit theories of intelligence: its relationships with self-efficacy, goal orientations, and self-regulated learning. *The International Journal of Learning*, 15(2), 47-56.
- Adams, W. K., Perkins, K. K., Podolefsky, N. S., Dubson, M., Finkelstein, N. D., & Wieman, C. E. (2006). New instrument for measuring student beliefs about physics and learning physics: The Colorado Learning Attitudes about Science Survey. *Physical Review Special Topics-Physics Education Research*, 2(1), 2-15.
- AERA, APA, & NCME. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Alexander, P. A., & Jetton, T. L. (1996). The role of importance and interest in the processing of text. *Educational Psychology Review*, 8(1), 89-121.
- Andrei, S., & Simons, E. A. (2008). A Multi-Institutional Assessment of the Use of POGIL in Organic Chemistry *Process Oriented Guided Inquiry Learning (POGIL)* (Vol. 994, pp. 226-239): American Chemical Society.

- Andrew, S. (1998). Self-efficacy as a predictor of academic performance in science. *Journal of Advancing Nursing, 27*(3), 596-603.
- APA, AERA, & NCME. (1954). *Technical recommendations for psychological tests and diagnostic techniques* (Vol. 51): American Psychological Association.
- Arjoon, J. A., Xu, X., & Lewis, J. E. (2013). Understanding the State of the Art for Measurement in Chemistry Education Research: Examining the Psychometric Evidence. *Journal of Chemical Education, 90*(5), 536-545.
- Asher, S. R., Hymel, S., & Wigfield, A. (1978). Influence of topic interest on children's reading comprehension. *Journal of Literacy Research, 10*(1), 35-47.
- Atkinson, J. W. (1957). Motivational determinants of risk-taking behavior. *Psychological Review, 64*(6, Pt. 1), 359-372.
- Atkinson, J. W. (1964). *An introduction to motivation*. Oxford, England: Van Nostrand.
- Aydin, Y., Uzuntiryaki, E., & Demirdöğ'en, B. (2011). Interplay of motivational and cognitive strategies in predicting self-efficacy and anxiety. *Educational Psychology, 31*(1), 55-66.
- Baldwin, J. A., Ebert-May, D., & Burns, D. J. (1999). The development of a college biology self-efficacy instrument for nonmajors. *Science Education, 83*(4), 397-408.
- Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191-215.

- Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist*, 37(2), 122-147.
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice Hall.
- Bandura, A. (1989). Human agency in social cognitive theory. *American Psychologist*, 44(9), 1175-1184.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.
- Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology*, 52(1), 1-26.
- Barbera, J., Adams, W. K., Wieman, C. E., & Perkins, K. K. (2008). Modifying and validating the Colorado Learning Attitudes about Science Survey for use in chemistry. *Journal of Chemical Education*, 85(10), 1435-1439.
- Barbera, J., & VandenPlas, J. R. (2011). All Assessment Materials Are Not Created Equal: The Myths about Instrument Development, Validity, and Reliability *Investigating Classroom Myths through Research on Teaching and Learning* (Vol. 1074, pp. 177-193): American Chemical Society.
- Barrows, H. S., & Tamblyn, R. M. (1980). *Problem-based learning: An approach to medical education*: Springer Publishing Company.
- Bathgate, M. E., Schunn, C. D., & Correnti, R. (2014). Children's Motivation Toward Science Across Contexts, Manner of Interaction, and Topic. *Science Education*, 98(2), 189-215.

- Bauer, C. F. (2005). Beyond "student attitudes": Chemistry self-concept inventory for assessment of the affective component of student learning. *Journal of Chemical Education*, 82(12), 1864-1870.
- Bauer, C. F. (2008). Attitude toward Chemistry: A Semantic Differential Instrument for Assessing Curriculum Impacts. *Journal of Chemical Education*, 85(10), 1440-1445.
- Belsey, D.A. (1991). *Conditioning diagnostics: Collinearity and weak data in regression*. New York: Wiley.
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107(2), 238-246.
- Benware, C. A., & Deci, E. L. (1984). Quality of learning with an active versus passive motivational set. *American Educational Research Journal*, 21(4), 755-765.
- Bernstein, M. R. (1955). Relationship between interest and reading comprehension. *The Journal of Educational Research*, 49(4), 283-288.
- Betz, N. E., & Hackett, G. (1983). The relationship of mathematics self-efficacy expectations to the selection of science-based college majors. *Journal of Vocational Behavior*, 23(3), 329-345.
- Black, A. E., & Deci, E. L. (2000). The effects of instructors' autonomy support and students' autonomous motivation on learning organic chemistry: A self-determination theory perspective. *Science Education*, 84(6), 740-756.

- Blackwell, L. (2002). *Psychological mediators of student achievement during the transition to junior high school: The role of implicit theories*. (Unpublished Doctoral Dissertation), Columbia University, New York.
- Blackwell, L., Trzesniewski, K. H., & Dweck, C. S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child Development, 78*(1), 246-263.
- Blalock, C. L., Lichtenstein, M. J., Owen, S., Pruski, L., Marshall, C., & Toepperwein, M.. (2008). In pursuit of validity: A comprehensive review of science attitude instruments 1935–2005. *International Journal of Science Education, 30*(7), 961-977.
- Bong, M. (2001). Role of self-efficacy and task-value in predicting college students' course performance and future enrollment intentions. *Contemporary Educational Psychology, 26*(4), 553-570.
- Bong, M., & Clark, R. E. (1999). Comparison between self-concept and self-efficacy in academic motivation research. *Educational Psychologist, 34*(3), 139-153.
- Born, W. K., Revelle, W., & Pinto, L. H. (2002). Improving biology performance with workshop groups. *Journal of Science Education and Technology, 11*(4), 347-365.
- Bouffard-Bouchard, T. (1990). Influence of self-efficacy on performance in a cognitive task. *The Journal of Social Psychology, 130*(3), 353-363.

- Bouffard-Bouchard, T., Parent, S., & Larivee, S.. (1991). Influence of self-efficacy on self-regulation and performance among junior and senior high-school age students. *International Journal of Behavioral Development, 14*(2), 153-164.
- Brandriet, A. R., Ward, R. M., & Bretz, S. L. (2013). Modeling meaningful learning in chemistry using structural equation modeling. *Chemistry Education Research and Practice, 14*(4), 421-430.
- Brophy, J. E. (2010). *Motivating students to learn*. New York: Routledge.
- Brown, S. D. (2010). A Process-Oriented Guided Inquiry Approach to Teaching Medicinal Chemistry. *American Journal of Pharmaceutical Education, 74*(7), 1-6.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford Press.
- Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research, 21*(2), 230-258.
- Burnette, J. L., O'Boyle, E. H., VanEpps, E. M., Pollack, J. M., & Finkel, E. J. (2013). Mind-sets matter: A meta-analytic review of implicit theories and self-regulation. *Psychological Bulletin, 139*(3), 655-701.
- Byrne, B. M. (2013). *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications, and programming*: Psychology Press.

- Campbell, N. K., & Hackett, G. (1986). The effects of mathematics task performance on math self-efficacy and task interest. *Journal of Vocational Behavior, 28*(2), 149-162.
- Carter, C. , LaRussa, M. A., & Bodner, G. M. (1987). A study of two measures of spatial ability as predictors of success in different levels of general chemistry. *Journal of Research in Science Teaching, 24*(7), 645-657.
- Chase, A., Pakhira, D., & Stains, M. (2013). Implementing Process-Oriented, Guided-Inquiry Learning for the First Time: Adaptations and Short-Term Impacts on Students' Attitude and Performance. *Journal of Chemical Education, 90*(4), 409-416.
- Chen, W., & Wong, Y. (2014). Chinese mindset: theories of intelligence, goal orientation and academic achievement in Hong Kong students. *Educational Psychology, 1-12*.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*(2), 233-255.
- Cohen, G. L., & G. J. (2008). Identity, Belonging, and Achievement: A Model, Interventions, Implications. *Current Directions in Psychological Science, 17*(6), 365-369.
- Cohen, G. L., Garcia, J., Apfel, N., & Master, A. (2006). Reducing the racial achievement gap: A social-psychological intervention. *Science, 313*(5791), 1307-1310.

- Cohen, G. L., Garcia, J., Purdie-Vaughns, V., Apfel, N., & Brzustoski, P. (2009). Recursive processes in self-affirmation: Intervening to close the minority achievement gap. *Science*, *324*(5925), 400-403.
- Cohen, J. (1992). Statistical power analysis. *Current Directions in Psychological Science*, *1*(3), 98-101.
- Cole, D. A., & Preacher, K. J. (2014). Manifest variable path analysis: Potentially serious and misleading consequences due to uncorrected measurement error. *Psychological Methods*, *19*(2), 300-315.
- Coleman, M. J., & Fults, B. A. (1982). Self-concept and the gifted classroom: The role of social comparisons. *Gifted Child Quarterly*, *26*, 116-120.
- Cooper, C. I., & Pearson, P. T. (2012). A Genetically Optimized Predictive System for Success in General Chemistry Using a Diagnostic Algebra Test. *Journal of Science Education and Technology*, *21*(1), 197-205.
- Cooper, M. (2010). The Case for Reform of the Undergraduate General Chemistry Curriculum. *Journal of Chemical Education*, *87*(3), 231-232.
- Covington, M. V., & Omelich, C. L. (1979). Effort: The double-edged sword in school achievement. *Journal of Educational Psychology*, *71*(2), 169-182.
- Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches*. Thousand Oaks, CA: Sage
- Creswell, J. W. (2013a). *Qualitative inquiry and research design: Choosing among five approaches* (3rd ed.). Thousand Oaks, CA: Sage.
- Creswell, J. W. (2013b). *Research design: Qualitative, quantitative, and mixed methods approaches*. Thousand Oaks, CA: Sage.

- Creswell, J. W., Plano-Clark, V. L., Gutmann, M. L., & Hanson, W. E. (2003).
Advanced mixed methods research designs. In A. Tashakkori & C. Teddlie
(Eds.), *Handbook of Mixed Methods and Behavioral Research* (pp. 209-
240). Thousand Oaks, CA: Sage.
- Crotty, M. (1998). *The foundations of social research: Meaning and perspective
in the research process*. London: Sage.
- Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics
to nonnormality and specification error in confirmatory factor analysis.
Psychological Methods, 1(1), 16-29.
- Dalgety, J., Coll, R. K., & Jones, A. (2003). Development of Chemistry Attitudes
and Experiences Questionnaire (CAEQ). *Journal of research in Science
Teaching, 40*(7), 649-668.
- Dalgety, J., & Coll, R. K. (2006a). Exploring first-year science students' chemistry
self-efficacy. *International Journal of Science and Mathematics Education,*
4(1), 97-116.
- Dalgety, J., & Coll, R. K. (2006b). The influence of first-year chemistry students'
learning experiences on their educational choices. *Assessment &
Evaluation in Higher Education, 31*(3), 303-328.
- Dalgety, J., & Salter, D. (2002). The development of the chemistry attitudes and
experiences questionnaire (CAEQ). *Chemical Education Research and
Practice, 3*(1), 19-32.

- Dancy, M., & Henderson, C. (2010). Pedagogical practices and instructional change of physics faculty. *American Journal of Physics*, 78(10), 1056-1063.
- Daubenmire, P. L., & Bunce, D. M. (2008). What Do Students Experience during POGIL Instruction? *Process Oriented Guided Inquiry Learning (POGIL)* (Vol. 994, pp. 87-99): American Chemical Society.
- Dawes, J. (2008). Do data characteristics change according to the number of scale points used? An experiment using 5 point, 7 point and 10 point scales. *International Journal of Market Research*, 51(1), 61-77.
- Deci, E. L. (1992). The Relation of Interest to the Motivation of Behavior: A Self-Determination Theory Perspective in *The role of interest in learning and development* (pp. 43-70): Lawrence Erlbaum Associates, Inc.
- Desimone, L. M., & Le Floch, K. C. (2004). Are we asking the right questions? Using cognitive interviews to improve surveys in education research. *Educational Evaluation and Policy Analysis*, 26(1), 1-22.
- Dewey, J. (1913). *Interest and effort in education*. New York: Houghton Mifflin.
- Dods, R. F. (1996). A Problem-Based Learning Design for Teaching Biochemistry. *Journal of Chemical Education*, 73(3), 225-228.
- Dougherty, R. C., Bowen, C. W., Berger, T., Rees, W., Mellon, E. K., & Pulliam, E. (1995). Cooperative Learning and Enhanced Communication: Effects on Student Performance, Retention, and Attitudes in General Chemistry. *Journal of Chemical Education*, 72(9), 793-797.

- Dowling, D. M. (1978). *The development of a mathematics confidence scale and its application in the study of confidence in Women college student*. (Unpublished doctoral dissertation), The Ohio State University.
- Duncan, T. G., & McKeachie, W. J. (2005). The Making of the Motivated Strategies for Learning Questionnaire. *Educational Psychologist*, 40(2), 117-128.
- Dweck, C. S. (1986). Motivational Processes Affecting Learning. *American Psychologist*, 41(10), 1040-1048.
- Dweck, C. S. (1996). Implicit theories as organizers of goals and behavior. In P. M. Gollwitzer & J. A. Bargh (Eds.), *The psychology of action: Linking cognition and motivation to behavior* (pp. 69-90). New York, NY: Guilford Press.
- Dweck, C. S. (2000). *Self-theories: Their role in motivation, personality, and development*. Psychology Press.
- Dweck, C. S. (2002). The development of ability conceptions. In J. Eccles (Ed.), *Development of achievement motivation* (pp. 57-88): Academic Press.
- Dweck, C. S. (2008). Can personality be changed? The role of beliefs in personality and change. *Current Directions in Psychological Science*, 17(6), 391-394.
- Dweck, C. S. (2012). Implicit Theories. In P. V. Lange, A. Kruglanski & T. Higgins (Eds.), *Handbook of Theories of Social Psychology* (Vol. 2, pp. 43-61). London: Sage.

- Dweck, C. S., & Leggett, E. L. (1988). A social cognitive approach to motivation and personality. *Psychological Review*, 95(2), 256-273.
- Dweck, C. S., Chiu, C., & Hong, Y. (1995). Implicit theories: Elaboration and extension of the model. *Psychological Inquiry*, 6(4), 322-333.
- Eccles, J., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J. L., & Midgley, C. (1983). Expectancies, values, and academic behaviors. In J. T. Spence (Ed.), *Achievement and achievement motivation* (pp. 75-146). San Francisco: W.H. Freeman.
- Eccles, J., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, 53(1), 109-132.
- Elliot, A. J. (1999). Approach and avoidance motivation and achievement goals. *Educational Psychologist*, 34(3), 169-189.
- Elliott, E. S., & Dweck, C. S. (1988). Goals - An approach to motivation and achievement. *Journal of Personality and Social Psychology*, 54(1), 5-12.
- Estes, T. H., & Vaughan, J. L. (1973). Reading interest and comprehension: Implications. *The Reading Teacher*, 149-153.
- Fairweather, J. (2008). *Linking evidence and promising practices in science, technology, engineering, and mathematics (STEM) undergraduate education*. Paper presented at the A status report. Paper prepared for National Research Council on Science Education Workshop on Evidence on Promising Practices in Undergraduate Science, Technology, Engineering, and Mathematics (STEM) Education The National Academies Washington, DC.

- Farrell, J. J., Moog, R. S., & Spencer, J. N. (1999). A Guided-Inquiry General Chemistry Course. *Journal of Chemical Education*, 76(4), 570-574.
- Fay, K., & Boyd, M. J. (2010). Eta-squared. In N. J. Salkind (Ed.), *Encyclopedia of research design* (Vol. 1, pp. 422-425). Thousand Oaks, CA: Sage.
- Ferrell, B., & Barbera, J. (2015). Analysis of students' self-efficacy, interest, and effort beliefs in general chemistry. *Chemistry Education Research and Practice*, 16(2), 318-337.
- Fidell, L. S., & Tabachnick, B. G. (2003). Preparatory Data Analysis. In I. B. Weiner, J. A. Schinka & W. F. Velicer (Eds.), *Handbook of Psychology - Research Methods in Psychology* (Vol. 2, pp. 115 - 142). Hoboken, NJ: John Wiley and Sons.
- Flowerday, T., & Schraw, G. (2003). Effect of choice on cognitive and affective engagement. *The Journal of Educational Research*, 96(4), 207-215.
- Flynn, A. B., & Biggs, R. (2011). The Development and Implementation of a Problem-Based Learning Format in a Fourth-Year Undergraduate Synthetic Organic and Medicinal Chemistry Laboratory Course. *Journal of Chemical Education*, 89(1), 52-57.
- Fransson, A. (1977). On qualitative differences in learning: IV—Effects of intrinsic motivation and extrinsic test anxiety on process and outcome. *British Journal of Educational Psychology*, 47(3), 244-257.
- Frenzel, A. C., Goetz, T., Pekrun, R., & Watt, H. (2010). Development of mathematics interest in adolescence: Influences of gender, family, and school context. *Journal of Research on Adolescence*, 20(2), 507-537.

- Furr, M. R., & Bacharach, V. R. (2013). *Psychometrics: an introduction* (2nd ed.). Thousand Oaks, CA: Sage.
- Gebru, M. T., Phelps, A. J., & Wulfsberg, G. (2012). Effect of clickers versus online homework on students' long-term retention of general chemistry course material. *Chemistry Education Research and Practice*, 13(3), 325-329.
- Geiger, L., Jones, L., & Karre, I. (2008). Transforming lecture halls with cooperative learning. *Chemists' Guide to Effective Teaching II*, 43-64.
- Gist, M. E., & Mitchell, T. R. (1992). Self-Efficacy: A Theoretical Analysis of Its Determinants and Malleability. *The Academy of Management Review*, 17(2), 183-211.
- Glynn, S. M., & Koballa, T. R. (2006). Motivation to learn in college science. In J. J. Mintzes & W. H. Leonard (Eds.), *Handbook of college science teaching* (pp. 25-32). Arlington, VA: National Science Teachers Association Press.
- Glynn, S. M., Taasobshirazi, G., & Brickman, P. (2009). Science Motivation Questionnaire: Construct Validation With Nonscience Majors. *Journal of Research in Science Teaching*, 46(2), 127-146.
- Gonzalez, A., & Paoloni, V. P. (2015). Perceived autonomy-support, expectancy, value, metacognitive strategies and performance in chemistry: a structural equation model in undergraduates. *Chemistry Education Research and Practice*, 16(3), 640-653.
- Goodman, S., Jaffer, T., Keresztesi, M., Mamdani, F., Mokgatle, D., Musariri, M., . . . Schlechter, A. (2011). An investigation of the relationship between

- students' motivation and academic performance as mediated by effort. *South African Journal of Psychology*, 41(3), 373-385.
- Gore, P. A. (2006). Academic self-efficacy as a predictor of college outcomes: Two incremental validity studies. *Journal of Career Assessment*, 14(1), 92-115.
- Gosser, D. K., & Roth, V. (1998). The workshop chemistry project: Peer-led team-learning. *Journal of Chemical Education*, 75(2), 185-187.
- Gosser, D., Roth, V., Gafney, L., Kampmeier, J., Stozak, V., Varma-Nelson, P., . . . Weiner, M. (1996). Workshop chemistry: Overcoming the barriers to student success. *The Chemical Educator*, 1(1), 1-17.
- Gray, C. D., & Kinnear, P. R. (2012). *IBM SPSS statistics 19 made simple*. New York: Psychology Press.
- Greene, J. C. (1985). Relationships among learning and attribution theory motivational variables. *American Educational Research Journal*, 22(1), 65-78.
- Guion, R. M., & Cranny, C. J. (1982). A note on concurrent and predictive validity designs. *Journal of Applied Psychology*, 67, 239-244.
- Gunderson, E. A., Gripshover, S. J., Romero, C., Dweck, C. S., Goldin-Meadow, S., & Levine, S. C. (2013). Parent praise to 1-to 3-year-olds predicts children's motivational frameworks 5 years later. *Child Development*, 84(5), 1526-1541.

- Gungor, A., Eryilmaz, A., & Fakioglu, T. (2007). The relationship of freshmen's physics achievement and their related affective characteristics. *Journal of Research in Science Teaching*, 44(8), 1036-1056.
- Haase, R. F., & Ellis, M. V. (1987). Multivariate analysis of variance. *Journal of Counseling Psychology*, 34(4), 404-413.
- Hackett, G. (1985). Role of mathematics self-efficacy in the choice of math-related majors of college women and men: A path analysis. *Journal of Counseling Psychology*, 32(1), 47-56.
- Hackett, G., & Betz, N. E. (1981). A self-efficacy approach to the career development of women. *Journal of Vocational Behavior*, 18(3), 326-339.
- Hackett, G., & Betz, N. E. (1989). An Exploration of the Mathematics Self-Efficacy/Mathematics Performance Correspondence. *Journal for Research in Mathematics Education*, 20(3), 261-273. doi: 10.2307/749515
- Hamby Towns, M., & Grant, E. R. (1997). "I believe I will go out of this class actually knowing something": Cooperative learning activities in physical chemistry. *Journal of Research in Science Teaching*, 34(8), 819-835.
- Hancock, G. R. (1999). A sequential Scheffé-type respecification procedure for controlling type I error in exploratory structural equation model modification. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(2), 158-168.
- Hanson, D., & Wolfskill, T. (2000). Process Workshops - A New Model for Instruction. *Journal of Chemical Education*, 77(1), 120-130.

- Harackiewicz, J. M., Barron, K. E., Tauer, J. M., & Elliot, A. J. (2002). Predicting success in college: A longitudinal study of achievement goals and ability measures as predictors of interest and performance from freshman year through graduation. *Journal of Educational Psychology, 94*(3), 562-575.
- Harackiewicz, J. M., Barron, K. E., Tauer, J. M., Carter, S. M., & Elliot, A. J. (2000). Short-term and long-term consequences of achievement goals: Predicting interest and performance over time. *Journal of Educational Psychology, 92*(2), 316-330.
- Harackiewicz, J. M., Durik, A. M., Barron, K. E., Linnenbrink-Garcia, L., & Tauer, J. M. (2008). The role of achievement goals in the development of interest: Reciprocal relations between achievement goals, interest, and performance. *Journal of Educational Psychology, 100*(1), 105-122.
- Harter, S. (1998). The development of self-representations. In W. Damon & N. Eisenberg (Eds.), *Handbook of Child Psychology* (pp. 553-617). Hoboken, NJ: John Wiley & Sons.
- Häussler, P., & Hoffmann, L. (2002). An intervention study to enhance girls' interest, self-concept, and achievement in physics classes. *Journal of Research in Science Teaching, 39*(9), 870-888.
- Henderson, C., & Dancy, M. H. (2011). *Increasing the impact and diffusion of STEM education innovations*. White paper commissioned for the forum on Characterizing the Impact and Diffusion of Engineering Education Innovations. Washington, DC.

- Henson, R. K. (2001). Understanding internal consistency reliability estimates: A conceptual primer on coefficient alpha. *Measurement and Evaluation in Counseling and Development, 34*, 177-189.
- Heredia, K., & Lewis, J. E. (2012). A Psychometric Evaluation of the Colorado Learning Attitudes about Science Survey for Use in Chemistry. *Journal of Chemical Education, 89*(4), 436-441.
- Hermida, R. (2015). The problem of allowing correlated errors in structural equation modeling: concerns and considerations. *Computational Methods in Social Sciences, 3*(1), 1-17.
- Herreid, C. F., & Schiller, N. A. (2013). Case studies and the flipped classroom. *Journal of College Science Teaching, 42*(5), 62-66.
- Hicks, R. W., & Bevsek, H. M. (2011). Utilizing Problem-Based Learning in Qualitative Analysis Lab Experiments. *Journal of Chemical Education, 89*(2), 254-257.
- Hidi, S. (1990). Interest and its contribution as a mental resource for learning. *Review of Educational Research, 60*(4), 549-571.
- Hidi, S., & Baird, W. (1986). Interestingness- A neglected variable in discourse processing. *Cognitive Science, 10*(2), 179-194.
- Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist, 41*(2), 111-127.
- Hinkin, T. R. (1995). A review of scale development practices in the study of organizations. *Journal of Management, 21*(5), 967-988.

- Hoaglin, D. C., & Iglewicz, B. (1987). Fine-tuning some resistant rules for outlier labeling. *Journal of the American Statistical Association*, 82(400), 1147-1149.
- Hoaglin, D. C., Iglewicz, B., & Tukey, J. W. (1986). Performance of some resistant rules for outlier labeling. *Journal of the American Statistical Association*, 81(396), 991-999.
- Hockings, S. C., DeAngelis, K. J., & Frey, R. F. (2008). Peer-Led Team Learning in General Chemistry: Implementation and Evaluation. *Journal of Chemical Education*, 85(7), 990-996.
- Hong, Y. Y., Chiu, C., & Dweck, C. S. (1995). Implicit theories of intelligence: Reconsidering the role of confidence in achievement motivation. In M. Kernis (Ed.), *Efficacy, agency, and self-esteem* (pp. 197-216). New York: Plenum.
- Hong, Y. Y., Chiu, C., Dweck, C. S., Lin, D. M., & Wan, W. (1999). Implicit theories, attributions, and coping: A meaning system approach. *Journal of Personality and Social Psychology*, 77(3), 588-599.
- Hooper, D., Coughlan, J., & Mullen, M. (2008). Structural equation modelling: guidelines for determining model fit. *Electronic Journal of Business Research Methods*, 6(1), 53-60.
- Hu, L. T., & Bentler, P. M. (1995). Evaluating model fit. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 76-99). London: Sage.

- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1-55.
- Huang, C. (2011). Self-concept and academic achievement: A meta-analysis of longitudinal relations. *Journal of School Psychology*, 49(5), 505-528. doi: <http://dx.doi.org/10.1016/j.jsp.2011.07.001>
- Huck, S. (2012). *Reading Statistics and Research* (6th ed.). Boston, MA: Pearson.
- Hughes, K. S. (2011). Peer-assisted learning strategies in human anatomy & physiology. *The American Biology Teacher*, 73(3), 144-147.
- Hull-Blanks, E., Kurpius, S. E., Befort, C., Sollenberger, S., Nicpon, M. F., & Huser, L. (2005). Career goals and retention-related factors among college freshmen. *Journal of Career Development*, 32(1), 16-30.
- Hulleman, C. S., Durik, A. M., Schweigert, S. B., & Harackiewicz, J. M. (2008). Task values, achievement goals, and interest: An integrative analysis. *Journal of Educational Psychology*, 100(2), 398-416.
- Hulleman, C. S, Godes, O., Hendricks, B. L., & Harackiewicz, J. M. (2010). Enhancing interest and performance with a utility value intervention. *Journal of Educational Psychology*, 102(4), 880-895.
- Hulleman, C. S, & Harackiewicz, J. M. (2009). Promoting interest and performance in high school science classes. *Science*, 326(5958), 1410-1412.

- Jones, B. D., Wilkins, J. L. M., Long, M. H., & Wang, F. H. (2012). Testing a motivational model of achievement: How students' mathematical beliefs and interests are related to their achievement. *European Journal of Psychology of Education, 27*(1), 1-20.
- Jöreskog, K. G., & Sörbom, D. (2006). LISREL 8.80 for Windows. *Computer Software. Lincolnwood, IL: Scientific Software International, Inc.*
- Kampmeier, J. A., Varma-Nelson, P., & Wedegaertner, D. K. (2000). Peer-led team learning: Organic chemistry. In N. J. Pienta, M. Cooper & T. Greenbowe (Eds.), *Chemists' Guide to Effective Teaching* (pp. 122-145). New Jersey: Prentice Hall.
- Kelly, O. C., & Finlayson, O. E. (2007). Providing solutions through problem-based learning for the undergraduate 1st year chemistry laboratory. *Chemistry Education Research and Practice, 8*(3), 347-361.
- Kelly, O. C., & Finlayson, O. E. (2009). A hurdle too high? Students' experience of a PBL laboratory module. *Chemistry Education Research and Practice, 10*(1), 42-52.
- Kenny, D. A. (1979). *Correlation and Causality*. New York: Wiley.
- Kerns, E. M. (1981). *Chemistry self-efficacy and its relationship to the avoidance of chemistry-related majors and careers*. (Unpublished honors thesis), Ohio State University, Columbus.

- Kim, T., & Schallert, D. L. (2014). Mediating effects of teacher enthusiasm and peer enthusiasm on students' interest in the college classroom. *Contemporary Educational Psychology, 39*(2), 134-144.
- Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York: Guilford Press.
- Knafl, K., Deatrick, J., Gallo, A., Holcombe, G., Bakitas, M., Dixon, J., & Grey, M. (2007). The analysis and interpretation of cognitive interviews for instrument development. *Research in Nursing & Health, 30*(2), 224-234.
- Kogut, L. S. (1997). Using cooperative learning to enhance performance in general chemistry. *Journal of Chemical Education, 74*(6), 720-722.
- Komarraju, M., & Nadler, D. (2013). Self-efficacy and academic achievement: Why do implicit beliefs, goals, and effort regulation matter? *Learning and Individual Differences, 25*(0), 67-72.
- Krapp, A. (2002). Structural and dynamic aspects of interest development: Theoretical considerations from an ontogenetic perspective. *Learning and Instruction, 12*(4), 383-409.
- Krapp, A., Hidi, S., & Renninger, K. A. (1992). Interest, Learning and Development *The role of interest in learning and development* (pp. 3-25): Lawrence Erlbaum Associates, Inc.
- Kray, L. J., & Haselhuhn, M. P. (2007). Implicit negotiation beliefs and performance: experimental and longitudinal evidence. *Journal of Personality and Social Psychology, 93*(1), 49-64.

- Landis, R., Edwards, B. D., & Cortina, J. (2009). Correlated residuals among items in the estimation of measurement models. In R. J. Vandenberg & C. E. Lance (Eds.), *Statistical and methodological myths and urban legends: Doctrine, verity and fable in the organizational and social sciences*. New York: Routledge.
- Larson, L. M., Stephen, A., Bonitz, V. S., & Wu, T. (2014). Predicting Science Achievement in India Role of Gender, Self-Efficacy, Interests, and Effort. *Journal of Career Assessment, 22*(1), 89-101.
- 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.
- Lee, W., Lee, M., & Bong, M. (2014). Testing interest and self-efficacy as predictors of academic self-regulation and achievement. *Contemporary Educational Psychology, 39*(2), 86-99.
- Lent, R. W., Brown, S. D., Brenner, B., Chopra, S. B., Davis, T., Talleyrand, R., & Suthakaran, V. (2001). The role of contextual supports and barriers in the choice of math/science educational options: A test of social cognitive hypotheses. *Journal of Counseling Psychology, 48*(4), 474-483.
- Lent, R. W., Brown, S. D., & Hackett, G. (1994). Toward a unifying social cognitive theory of career and academic interest, choice, and performance. *Journal of Vocational Behavior, 45*(1), 79-122.

- Lent, R. W., Brown, S. D., & Larkin, K. C. (1984). Relation of self-efficacy expectations to academic achievement and persistence. *Journal of Counseling Psychology, 31*(3), 356-362.
- Lent, R. W., Brown, S. D., & Larkin, K. C. (1986). Self-efficacy in the prediction of academic performance and perceived career options. *Journal of Counseling Psychology, 33*(3), 265-269.
- Lent, R. W., & Hackett, G. (1987). Career self-efficacy: Empirical status and future directions. *Journal of Vocational Behavior, 30*(3), 347-382.
- Lent, R. W., Sheu, H., Singley, D., Schmidt, J. A., Schmidt, L. C., & Gloster, C. S. (2008). Longitudinal relations of self-efficacy to outcome expectations, interests, and major choice goals in engineering students. *Journal of Vocational Behavior, 73*(2), 328-335.
- Lewis, S. E. (2011). Retention and Reform: An Evaluation of Peer-Led Team Learning. *Journal of Chemical Education, 88*(6), 703-707.
- Lewis, S. E., & Lewis, J. E. (2005). Departing from Lectures: An Evaluation of a Peer-Led Guided Inquiry Alternative. *Journal of Chemical Education, 82*(1), 135-139.
- Lewis, S. E., & Lewis, J. E. (2007). Predicting at-risk students in general chemistry: comparing formal thought to a general achievement measure. *Chemistry Education Research and Practice, 8*(1), 32-51.

- Lewis, S. E., Shaw, J. L., Heitz, J. O., & Webster, G. H. (2009). Attitude Counts: Self-Concept and Success in General Chemistry. *Journal of Chemical Education*, 86(6), 744-749.
- Lichtenstein, M. J., Owen, S. V., Blalock, C. L., Liu, Y., Ramirez, K. A., Pruski, L. A., . . . Toepperwein, M. A. (2008). Psychometric reevaluation of the scientific attitude inventory-revised (SAI-II). *Journal of Research in Science Teaching*, 45(5), 600-616.
- Lightsey, R. (1999). Albert Bandura and the Exercise of Self-Efficacy. *Journal of Cognitive Psychotherapy*, 13(2), 158-166.
- Linnenbrink-Garcia, L., Durik, A. M., Conley, A. M., Barron, K. E., Tauer, J. M., Karabenick, S. A., & Harackiewicz, J. M. (2010). Measuring Situational Interest in Academic Domains. *Educational and Psychological Measurement*, 70(4), 647-671.
- Linnenbrink-Garcia, L., Patall, E. A., & Messersmith, E. E. (2013). Antecedents and consequences of situational interest. *British Journal of Educational Psychology*, 83(4), 591-614.
- Lloyd, B. W., & Spencer, J. N. (1994). The Forum: New Directions for General Chemistry: Recommendations of the Task Force on the General Chemistry Curriculum. *Journal of Chemical Education*, 71(3), 206-209.
- Loevinger, J. (1957). Objective tests as instruments of psychological theory: Monograph Supplement 9. *Psychological Reports*, 3(3), 635-694.
- Lopez F. G., & Lent, R. W. (1992). Sources of mathematics self-efficacy in high school students. *The Career Development Quarterly*, 41(1), 3-12.

- MacArthur, J. R., & Jones, L. L. (2008). A review of literature reports of clickers applicable to college chemistry classrooms. *Chemistry Education Research and Practice*, 9(3), 187-195.
- MacCallum, R. C. (2003). 2001 Presidential address: Working with imperfect models. *Multivariate Behavioral Research*, 38(1), 113-139.
- Malgwi, C. A., Howe, M. A., & Burnaby, P. A. (2005). Influences on Students' Choice of College Major. *Journal of Education for Business*, 80(5), 275-282.
- Marsh, H. W. (1986). Verbal and math self-concepts: An internal/external frame of reference model. *American Educational Research Journal*, 23(1), 129-149.
- Marsh, H. W. (1992). Self Description Questionnaire (SDQ) II: A theoretical and empirical basis for the measurement of multiple dimensions of adolescent self-concept. A test manual and research monograph. New South Wales, Australia: University of Western Sydney, Faculty of Education.
- Marsh, H. W., & Bailey, M. (1991). Confirmatory factor analyses of multitrait-multimethod data: A comparison of alternative models. *Applied Psychological Measurement*, 15(1), 47-70.
- Marsh, H. W., & Yeung, A. S. (1998). Longitudinal structural equation models of academic self-concept and achievement: Gender differences in the development of math and English constructs. *American Educational Research Journal*, 35(4), 705-738.

- Martella, R. C., Nelson, J. R., Morgan, R. L., & Marchand-Martella, N. E. (2013). *Understanding and interpreting educational research*. New York: Guilford Press.
- Martin, D. P., & Randall, D. W. (2008). Using an ACS General Chemistry Exam to Compare Traditional and POGIL Instruction *Process Oriented Guided Inquiry Learning (POGIL)* (Vol. 994, pp. 240-247): American Chemical Society.
- Marx, M. H. (1958). Motivation. In C. W. Harris (Ed.), *Encyclopedia of Educational Research* (3rd ed., pp. 888-901). New York: Macmillan.
- Mason, R. L., Gunst, R. F., & Hess, J. L. (1989). *Statistical design and analysis of experiments: With applications to engineering and science*. New York: Wiley.
- McDonnell, C., O'Connor, C., & Seery, M. K. (2007). Developing practical chemistry skills by means of student-driven problem based learning mini-projects. *Chemistry Education Research and Practice*, 8(2), 130-139.
- McFate, C., & Olmsted, J. (1999). Assessing Student Preparation through Placement Tests. *Journal of Chemical Education*, 76(4), 562.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed., pp. 13-103). New York: MacMillan.
- Messick, S. (1995). Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741-749.

- Midgley, C., Maehr, M. L., Hruda, L. Z., Anderman, E., Anderman, L., Freeman, K. E., . . . Middleton, M. J. (2000). *Manual for the Patterns of Adaptive Learning Scales*. Ann Arbor: University of Michigan.
- Mitchell, M. (1993). Situational interest: Its multifaceted structure in the secondary school mathematics classroom. *Journal of Educational Psychology, 85*(3), 424-436. doi: 10.1037/0022-0663.85.3.424
- Mitchell, Y. D., Ippolito, J., & Lewis, S. E. (2012). Evaluating Peer-Led Team Learning across the two semester General Chemistry sequence. *Chemistry Education Research and Practice, 13*(3), 378-383.
- Miyake, A., Kost-Smith, L. E., Finkelstein, N. D., Pollock, S. J., Cohen, Geoffrey L, & Ito, T. A. (2010). Reducing the gender achievement gap in college science: A classroom study of values affirmation. *Science, 330*, 1234-1237.
- Motivation, [Def.1]. (n.d.). Oxford Dictionaries Online. Retrieved May 30, 2014, from http://www.oxforddictionaries.com/us/definition/american_english/motivation?q=motivation.
- Multon, K. D., Brown, S. D., & Lent, R. W. (1991). Relation of self-efficacy beliefs to academic outcomes: A meta-analytic investigation. *Journal of Counseling Psychology, 38*(1), 30-38.
- Murphy, K. R., & Davidshofer, C. O. (2005). *Psychological testing: Principles and applications*. New Jersey: Prentice Hall.

- Nameroff, T. J., & Busch, D. H. (2004). Exploring the Molecular Vision: Report from a SOCED Invitational Conference. *Journal of Chemical Education*, 81(2), 177-179.
- National Research Council. (2003). *BIO2010: Transforming Undergraduate Education for Future Research Biologists*. Washington, DC: The National Academies Press.
- Nauta, M. M., Kahn, J. H., Angell, J. W., & Cantarelli, E. A. (2002). Identifying the antecedent in the relation between career interests and self-efficacy: Is it one, the other, or both? *Journal of Counseling Psychology*, 49(3), 290-301.
- Nicholls, J. G. (1984). Achievement motivation - Conceptions of ability, subjective experience, task choice, and performance. *Psychological Review*, 91(3), 328-346.
- Nielsen, S. E., Scaffidi, J. P., & Yeziarski, E. J. (2014). Detecting Art Forgeries: A Problem-Based Raman Spectroscopy Lab. *Journal of Chemical Education*, 91(3), 446-450.
- Nieswandt, M. (2007). Student affect and conceptual understanding in learning chemistry. *Journal of Research in Science Teaching*, 44(7), 908-937.
- Osborne, J., & Waters, E. (2002). Four assumptions of multiple regression that researchers should always test. *Practical Assessment, Research & Evaluation*, 8(2), 1-9.

- Overton, T. L., & Randles, C. A. (2015). Beyond problem-based learning: using dynamic PBL in chemistry. *Chemistry Education Research and Practice*, 16, 251-259.
- Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research*, 66(4), 543-578.
- Pajares, F., Britner, S. L., & Valiante, G.. (2000). Relation between Achievement Goals and Self-Beliefs of Middle School Students in Writing and Science. *Contemporary Educational Psychology*, 25(4), 406-422.
- Pajares, F., & Kranzler, J. (1995). Self-efficacy beliefs and general mental ability in mathematical problem-solving. *Contemporary Educational Psychology*, 20(4), 426-443.
- Pajares, F., & Miller, D. M. (1994). Role of self-efficacy and self-concept beliefs in mathematical problem solving: A path analysis. *Journal of Educational Psychology*, 86(2), 193-203.
- Pajares, F., & Miller, D. M. (1995). Mathematics self-efficacy and mathematics performances: The need for specificity of assessment. *Journal of Counseling Psychology*, 42(2), 190-198.
- Pallant, J. (2010). *SPSS Survival Manual* (4th ed.). Australia: Allen & Unwin Book Publishers.
- Paulson, D. R. (1999). Active learning and cooperative learning in the organic chemistry lecture class. *Journal of Chemical Education*, 76(8), 1136-1140.
- Pedhazur, E. J. (1997). *Multiple regression in behavioral research: Explanation and prediction* (3rd ed.). Fort Worth: Harcourt Brace.

- Phillips, K. E. S., & Grose-Fifer, J. (2011). A Performance Enhanced Interactive Learning Workshop Model as a Supplement for Organic Chemistry Instruction. *Journal of College Science Teaching*, 40(3), 90-98.
- Piers, E. V. (2002). *The Piers-Harris children's self concept scale*. Los Angeles, CA: Western Psychological Services
- Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, 82(1), 33-40.
- Pintrich, P. R., Smith, D. A. F, García, T., & McKeachie, W. J. (1993). Reliability and predictive validity of the Motivated Strategies for Learning Questionnaire (MSLQ). *Educational and Psychological Measurement*, 53(3), 801-813.
- Plass, J. L., O'Keefe, P. A., Homer, B. D., Case, J., Hayward, E. O., Stein, M., & Perlin, K. (2013). The impact of individual, competitive, and collaborative mathematics game play on learning, performance, and motivation. *Journal of Educational Psychology*, 105(4), 1050-1066.
- Podsakoff, P. M., MacKenzie, S. B., Lee, J., & Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88(5), 879-903.
- President's Council of Advisors on Science and Technology. (2012). *Engage to excel: Producing one million additional college graduates with degrees in science, technology, engineering, and mathematics*. Washington, DC:

Executive Office of the President, President's Council of Advisors on Science and Technology.

- Price, D. S., & Brooks, D. W. (2012). Extensiveness and perceptions of lecture demonstrations in the high school chemistry classroom. *Chemistry Education Research and Practice*, 13(4), 420-427.
- Psycharis, Sarantos. (2013). Exploring the Effects of the Computational Experiment Approach to the Epistemic Beliefs, the Motivation, the Use of Modeling Indicators and Conceptual Understanding in Three Different Computational Learning Environments. *Journal of Education and Training Studies*, 1(1), 69-87.
- Pugh, K. J., Linnenbrink-Garcia, L., Koskey, K. L. K., Stewart, V. C., & Manzey, C. (2010). Motivation, learning, and transformative experience: A study of deep engagement in science. *Science Education*, 94(1), 1-28.
- Quitadamo, I. J., Brahler, C. J., & Crouch, G. J. (2009). Peer-Led Team Learning: A Prospective Method for Increasing Critical Thinking in Undergraduate Science Courses. *Science Educator*, 18(1), 29-39.
- Rajan, N., & Marcus, L. (2009). Student Attitudes and Learning Outcomes from Process Oriented Guided-Inquiry Learning (POGIL) Strategy in an Introductory Chemistry Course for Non-Science Majors: An Action Research Study. *The Chemical Educator*, 14, 85-93.
- Ram, P. (1999). Problem-Based Learning in Undergraduate Instruction. A Sophomore Chemistry Laboratory. *Journal of Chemical Education*, 76(8), 1122-1126.

- Rattan, A., Good, C., & Dweck, C. S. (2012). "It's ok — Not everyone can be good at math": Instructors with an entity theory comfort (and demotivate) students. *Journal of Experimental Social Psychology, 48*(3), 731-737.
- Raykov, T., & Marcoulides, G. A. (2000). *A First Course in Structural Equation Modeling*. Mahwah, N.J.: Lawrence Erlbaum Associates, Inc.
- Reinhardt, B. M. (1996). Factors Affecting Coefficient Alpha: A Mini Monte Carlo Study. In B. Thompson (Ed.), *Advances in social science methodology* (Vol. 4, pp. 3-20). Greenwich, CT: JAI Press.
- Renninger, K. A. (1992). Individual interest and development: Implications for theory and practice *The role of interest in learning and development* (pp. 361-395): Lawrence Erlbaum Associates, Inc.
- Renninger, K. A. (2000). Individual interest and its implications for understanding intrinsic motivation. In C. Sansone & J. Harackiewicz (Eds.), *Intrinsic and extrinsic motivation: The search for optimal motivation and performance* (pp. 373-404). San Diego, CA: Academic Press.
- Renninger, K. A., & Hidi, S. (2002). Student interest and achievement: Developmental issues raised by a case study. In A. Wigfield & J. Eccles (Eds.), *Development of achievement motivation* (pp. 175-191). San Diego, CA: Academic Press.
- Robbins, S. B., Lauver, K., Le, H., Davis, D., Langley, R., & Carlstrom, A. (2004). Do psychosocial and study skill factors predict college outcomes? A meta-analysis. *Psychological Bulletin, 130*(2), 261-288.

- Rogers, C. M, Smith, M. D, & Coleman, M. J. (1978). Social comparison in the classroom: the relationship between academic achievement and self-concept. *Journal of Educational Psychology, 70*(1), 50-57.
- Ruder, S. M., & Hunnicutt, S. S. (2008). POGIL in Chemistry Courses at a Large Urban University: A Case Study *Process Oriented Guided Inquiry Learning (POGIL)* (Vol. 994, pp. 133-147): American Chemical Society.
- Ryan, R. M, & Deci, E. L. (2000a). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology, 25*(1), 54-67.
- Ryan, R. M, & Deci, E. L. (2000b). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*(1), 68-78.
- Ryan, R. M, & Deci, E. L. (2000c). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*(1), 68-78.
- Salta, K., & Koulougliotis, D. (2015). Assessing motivation to learn chemistry: adaptation and validation of Science Motivation Questionnaire II with Greek secondary school students. *Chemistry Education Research and Practice, 16*(2), 237-250.
- Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye & C. C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 399-419). Thousand Oaks, CA: Sage Publications, Inc.

- Savery, J. R., & Duffy, T. M. (1995). Problem based learning: An instructional model and its constructivist framework. *Educational Technology, 35*(5), 31-38.
- Schaller, C. P., Graham, K. J., Johnson, B. J., Jakubowski, H. V., McKenna, A. G., McIntee, E. J., . . . Peterson, A. A. (2015). Chemical Structure and Properties: A Modified Atoms-First, One-Semester Introductory Chemistry Course. *Journal of Chemical Education, 92*(2), 237-246.
- Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online, 8*(2), 23-74.
- Schiefele, H. (1974). *Lernmotivation und Motivlernen (Motivation to learn and acquisition of motives)*: Ehrenwirth München.
- Schiefele, U. (1991). Interest, learning, and motivation. *Educational Psychologist, 26*(3-4), 299-323.
- Schiefele, U. (1992). Topic Interest and Levels of Text Comprehension *The role of interest in learning and development* (pp. 151-182): Lawrence Erlbaum Associates, Inc.
- Schiefele, U. (1999). Interest and learning from text. *Scientific Studies of Reading, 3*(3), 257-279.

- Schiefele, U., Krapp, A., & Winteler, A. (1992). Interest as a Predictor of Academic Achievement: A Meta-Analysis of Research *The role of interest in learning and development* (pp. 183-212): Lawrence Erlbaum Associates, Inc.
- Schmid, S., Youl, D. J., George, A. V., & Read, J. R. (2012). Effectiveness of a short, intense bridging course for scaffolding students commencing university-level study of chemistry. *International Journal of Science Education, 34*(8), 1211-1234.
- Schmidt, H. G. (1983). Problem-based learning: rationale and description *Medical Education, 17*(1), 11-16.
- Schraw, G., Flowerday, T., & Lehman, S. (2001). Increasing situational interest in the classroom. *Educational Psychology Review, 13*(3), 211-224.
- Schroeder, J. D., & Greenbowe, T. J. (2008). Implementing POGIL in the lecture and the Science Writing Heuristic in the laboratory-student perceptions and performance in undergraduate organic chemistry. *Chemistry Education Research and Practice, 9*(2), 149-156.
- Schunk, D. (1989). Self-efficacy and cognitive skill learning. In C. Ames & R. Ames (Eds.), *Research on motivation in education* (Vol. 3, Goals and Cognitions, pp. 13-44). San Diego: Academic.
- Schunk, D. (2008). *Learning theories: an educational perspective* (5th ed.). Upper Saddle River, NJ: Prentice Hall.

- Schunk, D., Hanson, Antoinette R., & Cox, P. D. (1987). Peer-model attributes and children's achievement behaviors. *Journal of Educational Psychology*, 79(1), 54-61.
- Schunk, D., & Pajares, F. (2002). The development of academic self-efficacy. In A. Wigfield & J. Eccles (Eds.), *Development of achievement motivation* (pp. 16-29). San Diego, CA: Academic Press.
- Schunk, D., & Pintrich, P. R. (2002). *Motivation in education: Theory, research, and applications*. Englewood Cliffs, NJ: Prentice-Hall.
- Scientific Software International. (n.d.) *The analysis of continuous data from multiple groups*. Retrieved from <http://www.ssicentral.com/lisrel/techdocs/Session7.pdf>
- Sears, R. R. (1937). Initiation of the repression sequence by experienced failure. *Journal of Experimental Psychology*, 20(6), 570-580.
- Seligman, M. E. P. (1975). *Helplessness: On depression, development, and death*. San Francisco: W.H. Freeman.
- Semsar, K., Knight, J. K., Birol, G., & Smith, M. K. (2011). The Colorado Learning Attitudes about Science Survey (CLASS) for use in biology. *CBE-Life Sciences Education*, 10(3), 268-278.
- Sevincer, A. T., Kluge, L., & Oettingen, G. (2014). Implicit theories and motivational focus: Desired future versus present reality. *Motivation and Emotion*, 38(1), 36-46.
- Seymour, E., & Hewitt, N. M. (1997). *Talking about leaving: Why undergraduates leave the sciences* (Vol. 12): Westview Press Boulder, CO.

- Shank, G. D. (2002). *Qualitative research: A personal skills approach*. Upper Saddle River, NJ: Prentice Hall.
- Shell, D. F., Murphy, C. C., & Bruning, R. H. (1989). Self-efficacy and outcome expectancy mechanisms in reading and writing achievement. *Journal of Educational Psychology, 81*(1), 91-100.
- Shively, R. L., & Ryan, C. S. (2013). Longitudinal changes in college math students' implicit theories of intelligence. *Social Psychology of Education, 16*(2), 241-256.
- Siegel, R. G., Galassi, J. P., & Ware, W. B. (1985). A comparison of two models for predicting mathematics performance: Social learning versus math aptitude–anxiety. *Journal of Counseling Psychology, 32*(4), 531-538.
- Silvia, P. J. (2001). Interest and interests: The psychology of constructive capriciousness. *Review of General Psychology, 5*(3), 270-290.
- Silvia, P. J. (2003). Self-efficacy and interest: Experimental studies of optimal incompetence. *Journal of Vocational Behavior, 62*(2), 237-249.
- Singh, K., Granville, M., & Dika, S. (2002). Mathematics and science achievement: Effects of motivation, interest, and academic engagement. *Journal of Educational Research, 95*(6), 323-332.
- Slavin, R. E. (1991). Synthesis of Research on Cooperative Learning. *Educational Leadership, 48*(5), 71-81.
- Smist, J. M. (1993). *General Chemistry and Self-Efficacy*. Paper presented at the Paper presented at the American Chemical Society national meeting, Chicago, IL.

- Smith, P. L., & Fouad, N. A. (1999). Subject-matter specificity of self-efficacy, outcome expectancies, interests, and goals: Implications for the social-cognitive model. *Journal of Counseling Psychology, 46*(4), 461-471.
- Sorich, L., & Dweck, C. S. (1997). *Reliability data for new scales measuring students' beliefs about effort and responses to failure*. Unpublished raw data. Columbia University.
- Spencer, H. E. (1996). Mathematical SAT test scores and college chemistry grades. *Journal of Chemical Education, 73*(12), 1150-1153.
- Spencer, J. N. (1999). New directions in teaching chemistry: A philosophical and pedagogical basis. *Journal of Chemical Education, 76*(4), 566-569.
- Steele, C. M. (1988). The psychology of self-affirmation: Sustaining the integrity of the self. *Advances in Experimental Social Psychology, 21*, 261-302.
- Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research, 25*(2), 173-180.
- Stipek, D., & Gralinski, J. H. (1996). Children's beliefs about intelligence and school performance. *Journal of Educational Psychology, 88*(3), 397-407.
- Streiner, D. L. (2005). Finding our way: an introduction to path analysis. *Canadian Journal of Psychiatry, 50*(2), 115-122.
- Streiner, D. L. (2006). Building a Better Model: An Introduction to Structural Equation Modelling. *Canadian Journal of Psychiatry, 51*(5), 317-324.
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics*. Boston: Pearson Education.

- Tempelaar, D. T., Rienties, B., Giesbers, B., & Gijsselaers, W. H. (2015). The Pivotal Role of Effort Beliefs in Mediating Implicit Theories of Intelligence and Achievement Goals and Academic Motivations. *Social Psychology of Education, 18*(1), 101-120.
- Tien, L., Roth, V., Kampmeier, J. A. (2002). Implementation of a peer- led team learning instructional approach in an undergraduate organic chemistry course. *Journal of research in science teaching, 39*(7), 606-632.
- Tolman, E. C. (1955). Principles of performance. *Psychological Review, 62*(5), 315-326.
- Tosun, C. (2013). Adaptation of Chemistry Motivation Questionnaire-II to Turkish: A Validity and Reliability Study. *Journal of Education Faculty, 15*(1), 173-202.
- Tuan, H. L., Chin, C. C., & Shieh, S. H. (2005). The development of a questionnaire to measure students' motivation towards science learning. *International Journal of Science Education, 27*(6), 639-654.
- Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38*(1), 1-10.
- Tukey, John W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.
- Usher, E. L., & Pajares, F. (2008). Sources of self-efficacy in school: Critical review of the literature and future directions. *Review of Educational Research, 78*(4), 751-796.

- Uzuntiryaki, E. (2008). Exploring the Sources of Turkish Pre-service Chemistry Teachers' Chemistry Self-efficacy Beliefs. *Australian Journal of Teacher Education*, 33(6), 12-28.
- Uzuntiryaki, E., & Aydin, Y. C. (2009). Development and Validation of Chemistry Self-Efficacy Scale for College Students. *Research in Science Education*, 39(4), 539-551.
- Uzuntiryaki, E., & Aydin, Y. C. (2013). Predicting Critical Thinking Skills of University Students through Metacognitive Self-Regulation Skills and Chemistry Self-Efficacy. *Educational Sciences: Theory & Practice*, 13(1), 666-670.
- Vacha-Haase, T., Kogan, L. R., & Thompson, B. (2000). Sample compositions and variabilities in published studies versus those in test manuals: Validity of score reliability inductions. *Educational and Psychological Measurement*, 60(4), 509-522.
- 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.
- Villafane, S. M., Garcia, C. A., & Lewis, J. E. (2014). Exploring diverse students' trends in chemistry self-efficacy throughout a semester of college-level preparatory chemistry. *Chemistry Education Research and Practice*, 15(2), 114-127.

- Vuong, M., Brown-Welty, S., & Tracz, S. (2010). The effects of self-efficacy on academic success of first-generation college sophomore students. *Journal of College Student Development, 51*(1), 50-64.
- Wagner, R., Kantor, P. T., & Piasta, S. (2010). Latent variable. In N. J. Salkind (Ed.), *Encyclopedia of research design* (Vol. 2, pp. 696-698). Thousand Oaks, CA: Sage.
- Wamser, C. C. (2006). Peer-Led Team Learning in Organic Chemistry: Effects on Student Performance, Success, and Persistence in the Course. *Journal of Chemical Education, 83*(10), 1562-1566.
- Weiner, B. (1969). Motivation. In R. L. Ebel (Ed.), *Encyclopedia of Educational Research* (4th ed., pp. 735-742). New York: Macmillan.
- Weiner, B. (1985). An attributional theory of achievement-motivation and emotion. *Psychological Review, 92*(4), 548-573.
- Weiner, B. (1990a). History of motivational research in education. *Journal of Educational Psychology, 82*(4), 616-622.
- Weiner, B. (1990b). Motivation. In M. C. Alkin (Ed.), *Encyclopedia of Educational Research* (6th ed., pp. 860-865). New York: Macmillan.
- Weiner, B., Frieze, I., Kukla, A., Reed, L., Rest, S., & Rosenbaum, R. M. (1971). *Perceiving the causes of success and failure*. Morristown, NJ: General Learning Press.
- Welch, G. W. (2010). Confirmatory factor analysis. In N. J. Salkind (Ed.), *Encyclopedia of Research Design* (Vol. 3, pp. 216-220). Thousand Oaks, CA: Sage.

- Wigfield, A., & Cambria, J. (2010). Students' achievement values, goal orientations, and interest: Definitions, development, and relations to achievement outcomes. *Developmental Review, 30*(1), 1-35.
- Wigfield, A., & Eccles, J. (2000). Expectancy–value theory of achievement motivation. *Contemporary Educational Psychology, 25*(1), 68-81.
- Wigfield, A., & Eccles, J. (2002a). The development of competence beliefs, expectancies for success, and achievement values from childhood through adolescence. In A. Wigfield & J. S. Eccles (Eds.), *Development of achievement motivation* (pp. 91-120): Academic Press.
- Wigfield, A., & Eccles, J. (2002b). Introduction. In A. Wigfield & J. S. Eccles (Eds.), *Development of achievement motivation* (pp. 1-10). San Diego, CA: Academic Press.
- Wigfield, A., Eccles, J., Yoon, K. S., Harold, R. D., Arbretton, A. J. A., Freedman-Doan, C., & Blumenfeld, P. C. (1997). Change in children's competence beliefs and subjective task values across the elementary school years: A 3-year study. *Journal of Educational Psychology, 89*(3), 451-469.
- Winkelmann, K., Baloga, M., Marcinkowski, T., Giannoulis, C., Anquandah, G., & Cohen, P. (2015). Improving Students' Inquiry Skills and Self-Efficacy through Research-Inspired Modules in the General Chemistry Laboratory. *Journal of Chemical Education, 92*(2), 247-255.
- Woodward, Arthur E., Weiner, Michael, & Gosser, David. (1993). Problem solving workshops in general chemistry. *Journal of Chemical Education, 70*(8), 651.

- Wright, S. L., Jenkins-Guarnieri, M. A., & Murdock, J. L. (2012). Career development among first-year college students: College self-efficacy, student persistence, and academic success. *Journal of Career Development, 292-310*.
- Wylie, R. C. (1979). *The self-concept: Theory and research on selected topics* (Vol. 2): University of Nebraska Press.
- Xu, X., Kim, E. S., & Lewis, J. E. (2016). Sex difference in spatial ability for college students and exploration of measurement invariance. *Learning and Individual Differences, 45*, 176-184.
- Xu, X., & Lewis, J. E. (2011). Refinement of a Chemistry Attitude Measure for College Students. *Journal of Chemical Education, 88(5)*, 561-568.
- Xu, X., Villafane, S. M., & Lewis, J. E. (2013). College students' attitudes toward chemistry, conceptual knowledge and achievement: structural equation model analysis. *Chemistry Education Research and Practice, 14(2)*, 188-200.
- Yeager, D. S., & Walton, G. M. (2011). Social-psychological interventions in education They're not magic. *Review of Educational Research, 81(2)*, 267-301.
- Yeziarski, E. J., Bauer, C. F., Hunnicutt, S. S., Hanson, David, Amaral, Katie E., & Schneider, James P. (2008). POGIL Implementation in Large Classes: Strategies for Planning, Teaching, and Management *Process Oriented Guided Inquiry Learning (POGIL)* (Vol. 994, pp. 60-71): American Chemical Society.

- Zeyer, A., Çetin-Dindar, A., Md Zain, A. N., Jurišević, M., Devetak, I., & Odermatt, F. (2013). Systemizing: A cross-cultural constant for motivation to learn science. *Journal of Research in Science Teaching, 50*(9), 1047-1067.
- Zimmerman, B. J. (2000). Self-efficacy: An essential motive to learn. *Contemporary Educational Psychology, 25*(1), 82-91.
- Zimmerman, B. J., Bandura, A., & Martinez-Pons, M. (1992). Self-motivation for academic attainment: The role of self-efficacy beliefs and personal goal setting. *American Educational Research Journal, 29*(3), 663-676.
- Zusho, A., Pintrich, P. R., & Coppola, B. (2003). Skill and will: The role of motivation and cognition in the learning of college chemistry. *International Journal of Science Education, 25*(9), 1081-1094.

APPENDIX A
INSTITUTIONAL REVIEW BOARD APPROVALS
AND INFORMED CONSENT FORMS

CONSENT FORM FOR HUMAN PARTICIPANTS IN RESEARCH

Project Title: An investigation of students' interest, effort beliefs, and self-efficacy in General Chemistry
 Researcher: Brent Ferrell, Doctoral student in the chemistry education program
 Phone number: (970) 351-1291 Email: brent.ferrell@unco.edu
 Research Advisor: Dr. Jack Barbera, Assistant Professor, Department of Chemistry and Biochemistry
 Phone Number: (970) 351-2545 Email: jack.barbera@unco.edu

You are being asked to outline, in your own words, your interest, effort beliefs, and confidence toward completing tasks in your general chemistry course. Following that, you will be asked to explain your reasoning for the answer choices you made on a questionnaire that you have taken previously. Each session will last approximately 30 minutes. Our goal is to understand your reasoning for the answer choices you made on the questionnaire.

We do not foresee any risk to you by participating in this study. You may feel anxious or frustrated by taking quizzes or tests, but we hope to minimize these feelings because the outcome of this interview has no connection with your general chemistry course or your final grade. There is no direct benefit to you as part of this study. If at any point during the interview you wish to no longer participate in this survey, you can withdraw without penalty or need for explanation.

Confidentiality will be maintained during the course of data collection and analysis. Signed consent forms will be stored separately from the data so that names cannot be linked to the information collected. Each participant shall have a random eight-digit code assigned to them for confidentiality and data analysis purposes. Electronic data will be stored on a password locked computer and only be accessible to the primary researchers. The consent forms will be stored for a three - year period. The office number where the data will be locked and stored is Ross 3695 (the chemical education office).

Questions: If you have any questions about the design or results of this study, or about the nature of your participation, you may ask now or at any time during the course of the data collection and subsequent analysis. You may also contact me or my advisor at the phone numbers indicated at the top of this form.

Participation is voluntary. You may decide NOT to participate in this study and if you do begin participation you may still decide to stop and withdraw at any time. Having read the above and having had an opportunity to ask any questions, please sign below if you would like to participate in this research. A copy of this form will be given to you to retain for future reference. If you have any concerns about your selection or treatment as a research participant, please contact the Sponsored Programs and Academic Research Center, Kepner Hall, University of Northern Colorado Greeley, CO 80639; 970-351-1907.

Print name _____

Participant's Signature

Date

Researcher's Signature

Date

CONSENT FORM FOR HUMAN PARTICIPANTS IN RESEARCH

Project Title: An investigation of students' interest, effort beliefs, and self-efficacy in General Chemistry
 Researcher: Brent Ferrell, Doctoral student in the chemistry education program
 Phone number: (970) 351-1291 Email: brent.ferrell@unco.edu
 Research Advisor: Dr. Jack Barbera, Assistant Professor, Department of Chemistry and Biochemistry
 Phone Number: (970) 351-2545 Email: jack.barbera@unco.edu

The primary goal of this research project is to evaluate a questionnaire, which is composed of items designed to measure interest, self-efficacy, and effort beliefs toward chemistry for first-semester general chemistry students. The items used on this assessment instrument have been used in previous research studies, but have been adapted to fit the context of a college chemistry course. As a result of the changes made, the validity and reliability of the instrument data must be re-evaluated. If educators and researchers wish to use this (or any) instrument to obtain data that informs curricular impacts or research questions, a thorough understanding of the instruments psychometric properties must be established. Data generated from this quantitative study will inform future qualitative studies to investigate individual item interpretation.

Any risk associated from participating in this study will be no different than what you may experience in a normal testing situation in a chemistry course. You may feel anxious or frustrated by taking quizzes or tests, but we hope to minimize these feelings because the outcome of taking this assessment has no connection with your evaluation in your general chemistry course or your final grade. If you decide to let your survey responses be used in this research, your participation will be confidential and will not affect your grade in the course, either positively or negatively.

Confidentiality will be maintained during the course of data collection and analysis. Signed consent forms will be stored separately from the data so that names cannot be linked to the information collected. Each participant shall have a random eight-digit code assigned to him or her for confidentiality and data analysis purposes.

I understand that by signing this consent form I am allowing my responses to this assessment instrument to be used in this research study.

Questions: If you have any questions about the design or results of this study, or about the nature of your participation, you may contact the researcher at any time by contacting the researchers using the phone numbers indicated at the top of this form.

Participation is voluntary. You may decide NOT to participate in this study and if you do begin participation you may still decide to stop and withdraw at any time. Having read the above and having had an opportunity to ask any questions, please sign below if you would like to participate in this research. A copy of this form will be given to you to retain for future reference. If you have any concerns about your selection or treatment as a research participant, please contact the Office of Sponsored Programs, Kepner Hall, University of Northern Colorado Greeley, CO 80639; 970-351-2161.

Print name _____

Participant's Signature

Date

Researcher's Signature

Date



Institutional Review Board

DATE: August 21, 2013

TO: Brent Ferrell
FROM: University of Northern Colorado (UNCO) IRB

PROJECT TITLE: [495865-2] An investigation of students' interest, effort beliefs, and self-efficacy in general chemistry

SUBMISSION TYPE: Amendment/Modification

ACTION: APPROVAL/VERIFICATION OF EXEMPT STATUS

DECISION DATE: August 20, 2013

Thank you for your submission of Amendment/Modification materials for this project. The University of Northern Colorado (UNCO) IRB approves this project and verifies its status as EXEMPT according to federal IRB regulations.

We will retain a copy of this correspondence within our records for a duration of 4 years.

If you have any questions, please contact Sherry May at 970-351-1910 or Sherry.May@unco.edu. Please include your project title and reference number in all correspondence with this committee.

This letter has been electronically signed in accordance with all applicable regulations, and a copy is retained within University of Northern Colorado (UNCO) IRB's records.

CONSENT FORM FOR HUMAN PARTICIPANTS IN RESEARCH

Project Title: The impact of directed connections between text, homework, quizzes, and exams on students' interest, effort beliefs, self-efficacy, and performance in General Chemistry

Researcher: Brent Ferrell, Doctoral student in the Chemistry Education program

Phone number: (970) 351-1291

Email: brent.ferrell@unco.edu

Research Advisor: Dr. Jack Barbera, Associate Professor, Department of Chemistry and Biochemistry

Phone Number: (970) 351-2545

Email: jack.barbera@unco.edu

The primary goal of this research project is to evaluate an alternative method of providing connections to students related to their textbook, homework, quizzes, and exams in a general chemistry setting. The researchers are specifically interested in how this will affect students' motivation and performance in introductory chemistry. Motivational aspects of students will be measured by using an instrument, which is composed of items designed to measure interest, self-efficacy, and effort beliefs toward chemistry for first-semester general chemistry students. The items used on this assessment instrument have been evaluated in a previous research study, and have been shown to exhibit strong psychometric properties in the context of an introductory chemistry class. If educators and researchers wish to make claims about students' motivation, a proper measurement tool must be used to obtain data that informs curricular impacts or research questions. Data generated from this quantitative study may inform future instructional strategies and delineation of feedback.

Any risk associated from participating in this study will be no different than what you may experience in a normal testing situation in a chemistry course. You may feel anxious or frustrated by completing the survey questions, but we hope to minimize these feelings. The outcome of taking this assessment has no connection with your evaluation in your general chemistry course or your final grade. If you allow us to use your survey responses and course grades in this research, your participation will be confidential and will not affect your grade in the course, either positively or negatively. Your instructor will not see your individual results on this survey or if you have declined to participate. Your instructor will be given a set of composite results for the class as a whole.

Confidentiality will be maintained during the course of data collection and analysis. Signed consent forms will be stored separately from the data so that names cannot be linked to the information collected. Each participant shall have a random four-digit code assigned to him or her for confidentiality and data analysis purposes. Electronic data will be stored on a password protected University computer. All data (both paper and electronic) will be destroyed after 3 years.

I understand that by signing this consent form I am allowing my responses to this assessment instrument and course grades to be used in this research study.

Questions: If you have any questions about the design or results of this study, or about the nature of your participation, you may contact the researcher at any time using the phone numbers indicated at the top of this form.

Participation is voluntary. You may decide NOT to participate in this study and if you do begin participation you may still decide to stop and withdraw at any time. Having read the above and having had an opportunity to ask any questions, please sign below if you would like to participate in this research. A copy of this form will be given to you to retain for future reference. If you have any concerns about your selection or treatment as a research participant, please contact the Office of Sponsored Programs, Kepner Hall, University of Northern Colorado Greeley, CO 80639; 970-351-2161.

Participant's name (please print) _____

Participant's Signature

Date

Researcher's Signature

Date

*Institutional Review Board*

DATE: June 18, 2014

TO: Brent Ferrell
FROM: University of Northern Colorado (UNCO) IRB

PROJECT TITLE: [615315-1] The impact of directed connections between text, homework, quizzes, and exams on students' interest, effort beliefs, self-efficacy, and performance in General Chemistry

SUBMISSION TYPE: New Project

ACTION: APPROVAL/VERIFICATION OF EXEMPT STATUS
DECISION DATE: June 17, 2014

Thank you for your submission of New Project materials for this project. The University of Northern Colorado (UNCO) IRB approves this project and verifies its status as EXEMPT according to federal IRB regulations.

Brent -

Hello and thank you for a thorough and clear IRB application for your research.

Best wishes with your participant recruitment and data collection. Please don't hesitate to contact me with any IRB-related questions or concerns.

Sincerely,

Dr. Megan Stellino, UNC IRB Co-Chair

We will retain a copy of this correspondence within our records for a duration of 4 years.

If you have any questions, please contact Sherry May at 970-351-1910 or Sherry.May@unco.edu. Please include your project title and reference number in all correspondence with this committee.

This letter has been electronically signed in accordance with all applicable regulations, and a copy is retained within University of Northern Colorado (UNCO) IRB's records.

APPENDIX B
ANNOUNCEMENTS FOR PARTICIPATION

Announcement to students for initial data collection for scales of self-efficacy, interest, and effort beliefs

Instructor Statement

Researchers at the University of Northern Colorado are evaluating a questionnaire to assess several factors that can have a strong influence on student performance and continued enrollment in chemistry. This questionnaire has been designed to help instructors understand these factors more deeply with the hope of improving instruction in future chemistry courses.

Today you will be asked to take a 25-item questionnaire, along with seven demographic items that will help the researchers understand more about your individual background. Everyone will take the questionnaire, but your responses will only be used in this research study if you sign the consent form on the first page of the questionnaire.

Your responses will help in evaluating the quality of each question and the questionnaire as a whole. Your participation in the research will be kept confidential and your responses will not be used in any type of evaluation for this course. I will not see your individual responses or if you chose to allow the researchers to use your data. The researchers will only give me a summary report for the class.

If you are interested, the researchers will be conducting interviews regarding the questionnaire throughout the semester. Focused student feedback is very important in this type of study to design the most effective questionnaire. The interviews will last approximately 30 minutes. If you answer YES on the final question of the demographics form, and you are selected, you will be contacted via your university e-mail.

Please fill in your university e-mail and bubble in the letters and numbers where it says PDID. Please print and bubble your last name as well. Please do not write on the questionnaire itself, only the Scantron.

Without your help, this study would not be possible. Your participation is greatly appreciated! I will now pass out the questionnaire; you will have 15 minutes to complete the 25 questions.

TA Notes:

- Please make sure each student has put their PDID and last name on the Scantron and filled in the bubbles.
- Please make sure students have signed the consent form, if they wish to participate.
- Remind students to use a #2 pencil
- Remind students that they can take a colored consent form if they wish.

Researcher Statement prior to administration of revised scales of self-efficacy,
interest, and effort beliefs

I am a graduate student researcher here at the University of Northern Colorado and my dissertation project is to evaluate several factors that can influence your performance in this course. Part of my research involves your responses to a questionnaire I have been working on. This questionnaire has been designed to help instructors understand these important learning factors more deeply with the hope of improving instruction in future chemistry courses.

Today, I will give you a 19-item questionnaire, along with seven items that will help me understand more about your individual background. I am asking that everyone take the questionnaire, however, your responses will only be used in this research study if you sign the consent form on the first page of the questionnaire. Please note that by signing the consent form, you are also granting me access to your course grades at the end of the semester. It is vital to the study that I can use your course grades, but I will not see any of your grades until your instructor finalizes them at the end of the semester.

Your participation in the research will be kept completely confidential and your responses will not be used in any type of evaluation for this course. If you choose to allow me to use your responses in the study, I will be the only person to see your individual responses. Your instructor will never see what your individual responses are. Hence, this study has no bearing on your grade in this course.

Having said that, without your help, this study would not be possible. Your participation is greatly appreciated!

I will now pass out the questionnaire; you will have 15 minutes to complete the 19 questions.

Announcement to students prior to administration of revised scales of self-efficacy, interest, and effort beliefs

Instructor Statement

Researchers at the University of Northern Colorado are evaluating a questionnaire to assess several factors that can have a strong influence on student performance and continued enrollment in chemistry. This questionnaire has been designed to help instructors understand these factors more deeply with the hope of improving instruction in future chemistry courses.

Today you will be asked to take a 19-item questionnaire. Everyone will take the questionnaire, but your responses will only be used in this research study if you give consent to have them used with question 1 on the survey.

Please fill in your last name and bubble in the letters below on the Scantron.

Without your help, this study would not be possible. Your participation is greatly appreciated!

I will now pass out the questionnaire; you will have 15 minutes to complete the 19 questions.

TA Notes:

- Please make sure each student has put their last name on the Scantron and filled in the bubbles.
- Remind students to use a #2 pencil

Researcher Statement prior to administration of revised scales of effort beliefs

I am a graduate student researcher here at the University of Northern Colorado and my dissertation project is to evaluate several factors that can influence your performance in this course. Part of my research involves your responses to a questionnaire I have been working on. This questionnaire has been designed to help instructors understand these important learning factors more deeply with the hope of improving instruction in future chemistry courses.

Today, I will give you a 6-item questionnaire, along with three items that will help me understand more about your individual background. I am asking that everyone take the questionnaire, however, your responses will only be used in this research study if you agree to the consent form on the back page of the questionnaire. By answering "Agree" to question 1, you are allowing me to use your data for my research.

Your participation in the research will be kept completely confidential and your responses will not be used in any type of evaluation for this course. If you choose to allow me to use your responses in the study, I will be the only person to see your individual responses. Your instructor will never see what your individual responses are. Hence, this study has no bearing on your grade in this course.

Having said that, without your help, this study would not be possible. Your participation is greatly appreciated!

I will now pass out the questionnaire

APPENDIX C
SUPPORTING INFORMATION FOR
QUANTITATIVE STUDIES

Scales and Demographic Items

Initial Interest Scale

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
A	B	C	D	E
II 1. I am fascinated by chemistry.				A B C D E
II 2. I chose to take general chemistry because I'm really interested in the topic.				A B C D E
II 3. I am really excited about taking this class.				A B C D E
II 4. I am really looking forward to learning more about chemistry.				A B C D E
II 5. I think the field of chemistry is an important discipline.				A B C D E
II 6. I think that what we will study in General Chemistry will be important for me to know.				A B C D E
II 7. I think that what we will study in General Chemistry will be worthwhile for me to know.				A B C D E

Maintained Interest Scale

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
A	B	C	D	E
MI 1. What we are learning in chemistry class this semester is fascinating to me.				A B C D E
MI 2. This semester, I really enjoy the chemistry material we cover in class.				A B C D E
MI 3. I am excited about what we are learning in chemistry class this semester.				A B C D E
MI R4. To be honest, I don't find the chemistry material we cover in class interesting.				A B C D E
MI 5. What we are studying in chemistry class is useful for me to know.				A B C D E
MI 6. The things we are studying in chemistry this semester are important to me.				A B C D E
MI 7. What we are learning in chemistry this semester is important for my future goals.				A B C D E
MI 8. What we are learning in chemistry this semester can be applied to real life.				A B C D E

Self-Efficacy Scale

Very Poorly	Poorly	Average	Well	Very Well
A	B	C	D	E

SE 1. To what extent can you explain chemical laws and theories?	A	B	C	D	E
SE 2. How well can you choose an appropriate formula to solve a chemistry problem?	A	B	C	D	E
SE 3. How well can you describe the structure of an atom?	A	B	C	D	E
SE 4. How well can you describe the properties of elements by using the periodic table?	A	B	C	D	E
SE 5. How well can you read the formulas of elements and compounds?	A	B	C	D	E
SE 6. How well can you interpret chemical equations?	A	B	C	D	E
SE 7. How well can you interpret graphs/charts related to chemistry?	A	B	C	D	E
SE 8. How well can you solve chemistry problems?	A	B	C	D	E

Effort Beliefs Scale

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
A	B	C	D	E

EB 1R*. To tell the truth, when I work hard at chemistry, it makes me feel like I'm not very smart	A	B	C	D	E
EB 2R*. It doesn't matter how hard you work if you're not smart in chemistry, you won't do well in it.	A	B	C	D	E
EB 3R*. If you're not good at chemistry, working hard won't make you good at it.	A	B	C	D	E
EB 4R*. If chemistry is hard for someone, it means that he or she probably won't be able to do really well at it.	A	B	C	D	E
EB 5R*. If you're not doing well at chemistry, it's better to try something easier.	A	B	C	D	E
EB 6. When chemistry is hard, it just makes me want to work more on it, not less.	A	B	C	D	E
EB 7. If you don't work hard at chemistry and put in a lot of effort, you probably won't do well.	A	B	C	D	E
EB 8. The harder you work at chemistry, the better you will be at it.	A	B	C	D	E
EB 9. If a chemistry assignment is hard, it means I'll probably learn a lot doing it.	A	B	C	D	E

*R indicates item must be reverse-coded prior to analysis

Demographics Form 1

This information is important to help us understand more about your individual backgrounds. Please mark your responses using the appropriate letter for each choice on the Scantron sheet. Please do not mark on this sheet. Thank you!!

D1. Gender : **A**- Male **B** - Female

D2. Age : **A** - <18 **B** – 18-20 **C** – 21-23 **D** – 24-25 **E** - >25

D3. How many years have you been in college?

A – This is my first semester **B** – 1 yr. **C** – 2 yrs. **D** – 3 yrs. **E** - >3 yrs.

D4. Is this your first chemistry class in college? **A** - Yes **B** - No

D5. How long ago did you take high school chemistry?

A – I did not take chemistry in high school

B – 1 yr. ago **C** – 2 yrs. ago **D** – 3 yrs. ago **E** - > 3 yrs. ago

D6. What is your declared major?

A – Chemistry (including Forensics, Biochemistry, Teaching, or Pre-Health)

B – Other Science (Biology, Physics, or Mathematics)

C – Other (including Sports & Exercise Science, Nursing, Earth Science, Statistics)

D – Undeclared

D7. Would you be willing to participate in a 30-minute interview regarding your interest, effort beliefs, and self-efficacy about chemistry? These interviews help us to further understand your responses and how the items make sense to you so we can make improvements to the questionnaire.

A - YES B - NO

Demographics Form 2

This information is important to help us understand more about your individual backgrounds. Please mark your responses using the appropriate letter for each choice on the Scantron sheet. Please do not mark on this sheet. Thank you!!

D1. Gender : **A**- Male **B** - Female

D2. Age : **A** - <18 **B** – 18-20 **C** – 21-23 **D** – 24-25 **E** - >25

D3. Ethnicity: **A** – African American **B** – Caucasian (White) **C** - Hispanic American
 D – Asian American **E** - Other

D4. How many years have you been in college?

A – This is my first semester **B** – 1 yr. **C** – 2 yrs. **D** – 3 yrs. **E** - >3 yrs.

D5. Is this your first chemistry class in college? **A** - Yes **B** - No

D6. How long ago did you take high school chemistry?

A – I did not take chemistry in high school

B – 1 yr. ago **C** – 2 yrs. ago **D** – 3 yrs. ago **E** - > 3 yrs. ago

D7. What is your declared major?

A – Chemistry (including Forensics, Biochemistry, Teaching, or Pre-Health)

B – Other Science (Biology, Physics, or Mathematics)

C – Other (including Sports & Exercise Science, Nursing, Earth Science, Statistics)

D – Undeclared

Table C1. Time 1 item-level descriptive statistics ($n = 373$)

Item	Mean	Std. Dev.	Skew	Kurtosis
II1	3.59	0.95	-0.31	-0.33
II2	3.09	1.07	0.06	-0.68
II3	3.40	0.94	-0.10	-0.23
II4	3.81	0.86	-0.67	0.46
II5	3.97	0.79	-0.73	1.14
II6	4.11	0.88	-1.20	1.82
II7	4.09	0.81	-0.98	1.51
EB1R	3.21	1.15	-0.24	-0.77
EB2R	3.96	0.92	-0.86	0.45
EB3R	4.19	0.79	-1.15	2.07
EB4R	3.92	0.87	-0.90	0.76
EB5R	3.87	0.78	-0.44	-0.04
EB6	3.48	1.03	-0.52	-0.34
EB7	4.07	0.80	-1.10	1.98
EB8	4.22	0.76	-1.36	3.54
EB9	3.57	0.84	-0.25	-0.10
SE1	2.47	0.86	-0.04	-0.53
SE2	2.63	0.93	-0.04	-0.53
SE3	3.14	1.04	-0.01	-0.53
SE4	3.16	0.95	-0.15	-0.45
SE5	3.12	0.95	-0.04	-0.21
SE6	2.91	0.92	-0.08	-0.23
SE7	3.16	0.86	-0.14	0.16
SE8	2.94	0.89	-0.26	-0.14

II – Initial interest, EB – Effort beliefs, SE – Self-efficacy, R – indicates item has been reverse-coded

Table C2. Time 2 descriptive statistics for maintained interest items ($n = 294$)

Item	Mean	Std. Dev.	Skew	Kurtosis
MI 1	3.32	.97	-.26	-.16
MI 2	3.23	.96	-.11	-.36
MI 3	3.15	.93	-.02	-.16
MI R4	3.20	1.15	-.24	-.96
MI 5	3.73	.96	-.50	-.23
MI 6	3.37	.97	-.18	-.29
MI 7	3.75	1.07	-.67	-.28
MI 8	3.56	.94	-.23	-.29

MFeel – Maintained interest (feeling), MVal – Maintained interest (value),
R – indicates item has been reverse-coded

Table C3. Mean scores and differences between chemistry majors and non-science majors on interest and self-efficacy scales for the cross-validation study sub-sample

Scale	Chemistry majors $n = 23$ Mean score (SD)	Non-science majors $n = 109$ Mean score (SD)	Mean difference ^a (effect size ^b)	p -value
Initial interest (feeling)	3.88 (0.75)	2.98 (0.83)	0.90 (0.70)	< 0.001
Initial interest (value)	4.50 (0.52)	3.67 (1.07)	0.83 (0.56)	<0.001
Maintained interest (feeling)	3.80 (0.85)	2.99 (0.92)	0.81 (0.56)	<0.001
Maintained interest (value)	3.85 (0.80)	3.30 (0.82)	0.55 (0.45)	0.004

^aBased on planned contrasts ^bEffect size represented by *Cohen's d* – small (0.20), medium (0.50), large (0.80) (Cohen, 1992)

CFA model diagrams

The figures below show the revised models for each scale for each time point. In the following figures, indicators are represented with a boxed border and latent variables are represented with an oval border. Error terms are represented by arrows pointing toward the indicators. The factor loadings are the numbers between the latent variables and the indicators. Correlations between latent variables are indicated by the number next to the double arrows in between two factors.

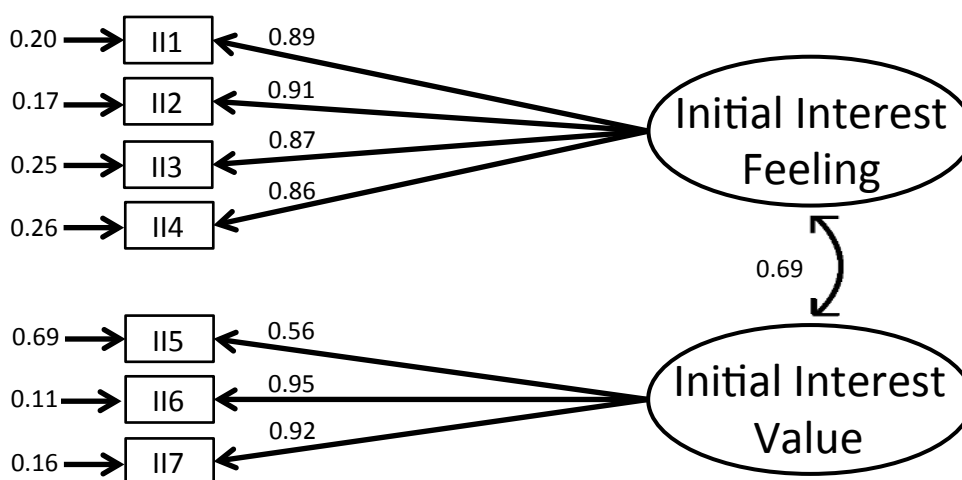


Figure C1. Time 1 initial interest CFA model (n = 373)

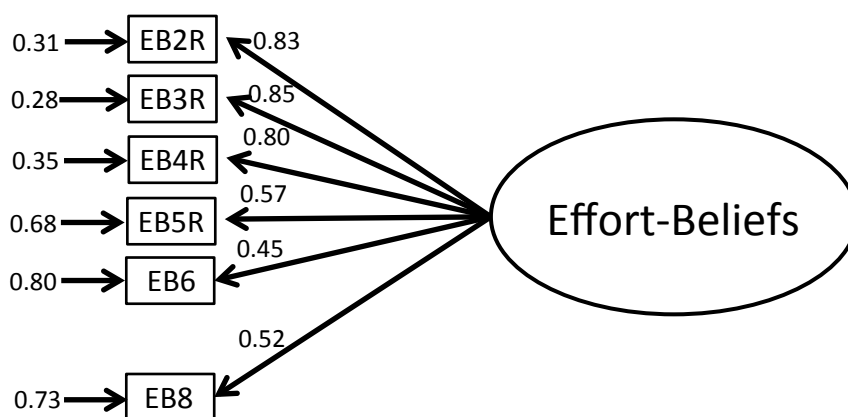


Figure C2. Time 1 effort beliefs CFA model (n = 373)

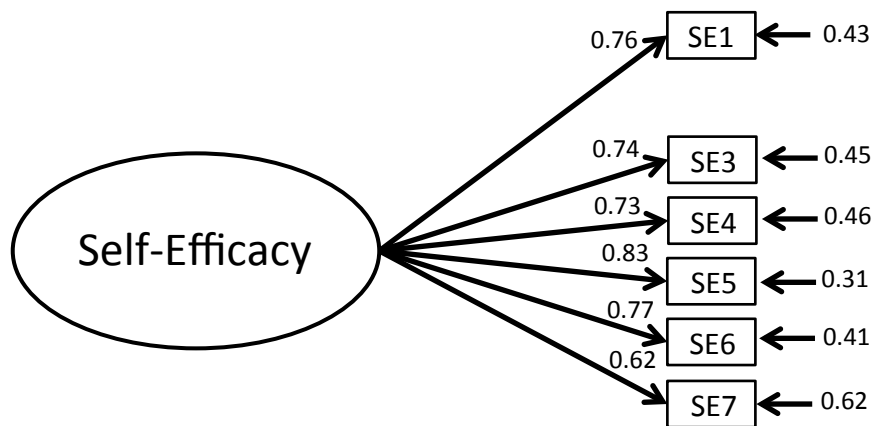


Figure C3. Time 1 self-efficacy CFA model (n = 373)

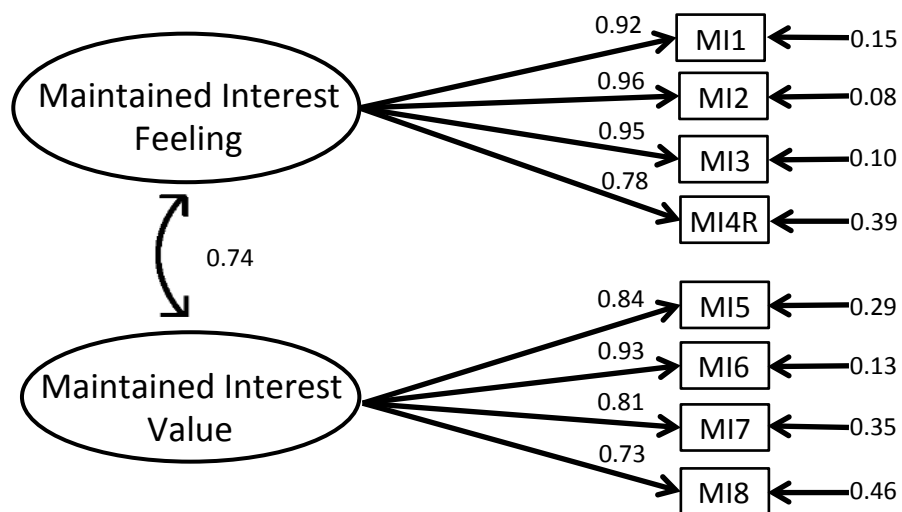


Figure C4. Time 2 maintained interest CFA model (n = 294)

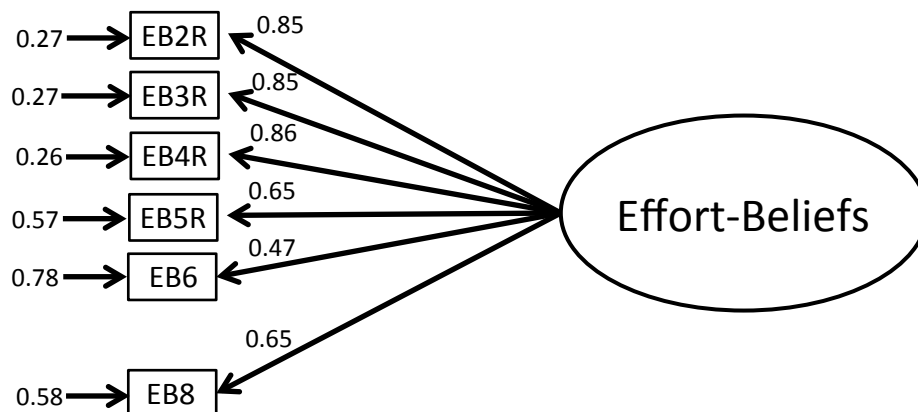


Figure C5. Time 2 effort beliefs CFA model (n = 294)

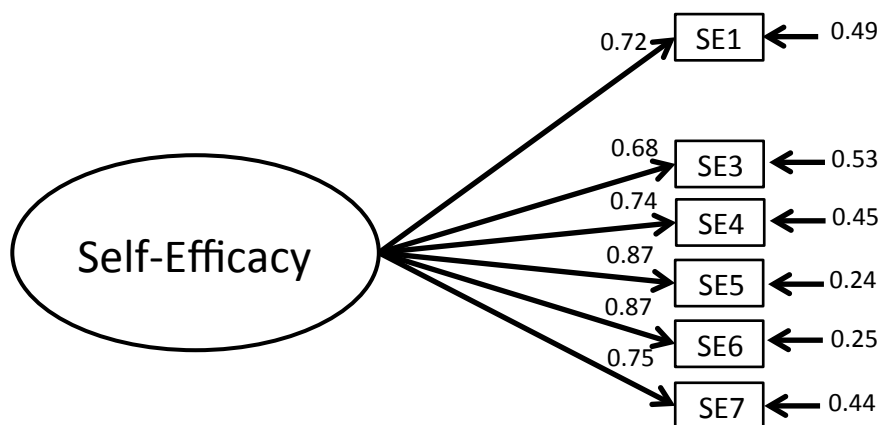


Figure C6. Time 2 self-efficacy CFA model (n = 294)

Table C4. Planned contrast results for both treatment sections versus the control section from MANCOVA analysis

Scale	Contrast estimate	Standard error	<i>p</i> - value
Maintained interest - feel	0.657	0.149	< 0.001
Effort beliefs T2	0.356	0.114	0.002
Self-efficacy T2	0.292	0.098	0.003

Note: A positive contrast estimate indicates the linear combination of the mean scores for the treatment sections was higher than the mean score of the control section

Table C5. Planned contrast results for values treatment versus study skills treatment from MANCOVA analysis

Scale	Contrast estimate	Standard error	<i>p</i> - value
Maintained interest - feel	0.087	0.175	0.618
Effort beliefs T2	0.062	0.134	0.647
Self-efficacy T2	0.125	0.115	0.279

Note: A positive contrast estimate indicates the linear combination of the mean scores for the treatment sections was higher than the mean score of the control section

APPENDIX D

**TIME TABLE AND SUMMARY FOR
ALL INTERVENTIONS**

Data Collection / Course Modifications

Time	Values Intervention	Study Skills
Day 1 (Aug 25)	Collect data with Initial Interest, Self-Efficacy, and Effort Beliefs Scales	Collect data with Initial Interest, Self-Efficacy, and Effort Beliefs Scales
Quiz 1 (Aug 29)	Values Affirmation Writing Assignment	Past Science/Math Experience Writing Assignment
Week 3 (Sep 9)	Intelligence Lecture (~20 minutes)	Study Skills Lecture (~20 minutes)
Pre-Exam 2 (Oct 7)	Collect data with Self-Efficacy and Effort Beliefs Scales and Values Affirmation Writing Assignment	Collect data with Self-Efficacy and Effort Beliefs Scales and Explaining a Concept Writing Assignment
Pre-Exam 4 (week 13)	Collect data with Maintained Interest, Self-Efficacy, and Effort Beliefs Scales	Collect data with Maintained Interest, Self-Efficacy, and Effort Beliefs Scales
Post Semester	Course % Score	Course % Score

Name (VALUES INTERVENTION GROUP - A)

Chem 111 Quiz 1

Brief writing assignment.

Effective communication is an important skill for success in all science-related fields. During class and in lab you regularly get practice with your oral communication skills; this quiz will allow you to begin practicing your written skills.

Your score is based on COMPLETION of the assignment not on the quality of your writing.

Part 1 – From the list of values below, circle the two or three that are **MOST IMPORTANT** to you.

Being good at art

Relationships with friends and family

Creativity

Government or politics

Independence

Learning and gaining knowledge

Athletic ability

Belonging to a social group (such as your community, racial group, or school club)

Music

Career

Spiritual or religious values

Part 3 – List the top two reasons why the values you selected are important to you.

Part 4 – Consider the values you selected. Using the options below please circle your level of agreement to the following statements.

i) These values have influenced my life.

Strongly Agree Agree Disagree Strongly Disagree

ii) In general, I try to live up to these values.

Strongly Agree Agree Disagree Strongly Disagree

iii) These values are an important part of who I am.

Strongly Agree Agree Disagree Strongly Disagree

iv) I care about these values.

Strongly Agree Agree Disagree Strongly Disagree

Name (STUDY SKILLS GROUP - B)

Chem 111 Quiz 1

Brief writing assignment.

Effective communication is an important skill for success in all science-related fields. During class and in lab you regularly get practice with your oral communication skills; this quiz will allow you to begin practicing your written skills.

Your score is based on COMPLETION of the assignment not on the quality of your writing.

Part 1 – From the list of values below, circle the two or three that are **LEAST IMPORTANT** to you.

Being good at art

Relationships with friends and family

Creativity

Government or politics

Independence

Learning and gaining knowledge

Athletic ability

Belonging to a social group (such as your community, racial group, or school club)

Music

Career

Spiritual or religious values

Part 3 – List the top two reasons why you think the values you selected might be important to someone else.

Part 4 – Consider the values you selected. Using the options below please circle your level of agreement to the following statements.

i) These values have influenced some people.

Strongly Agree **Agree** **Disagree** **Strongly Disagree**

ii) In general, some people try to live up to these values.

Strongly Agree **Agree** **Disagree** **Strongly Disagree**

iii) These values are an important part of who other people are.

Strongly Agree **Agree** **Disagree** **Strongly Disagree**

iv) Other people care about these values.

Strongly Agree **Agree** **Disagree** **Strongly Disagree**

Week 3 Lecture Outlines

Growth Mindset Lecture (VALUES INTERVENTION GROUP - A)

- 1) (~6 min) Complete assignment (see below)
- 2) (~2 min) Brief personal story about instructor's academic trajectory and amount of hard work and effort required to reach goals.
- 3) (~12 min) Watch TEDx talk by Eduardo Briceno (up to 7min 40sec), "The Power of Belief – Mindset and Success". After watching talk discuss the following questions:
 - 1) Can you change your mindset?
 - 2) How do you think you would go about making this change?

Problem Solving Skills (STUDY SKILLS GROUP - B)

- 1) (~6 min) Complete assignment (see below)
- 2) (~2 min) Brief personal story about instructor's problem solving skill development after facing course difficulties.
- 3) (~12 min) Poll class about how they study (or plan to study) for the course, write down methods on board. Discuss and expand upon the proper use of each method listed and fill in any major missing options.

Chem 111 – Week 3 In Class Assignment Name (VALUES INTERVENTION GROUP - A)

1) Write down one subject or activity that you do well (e.g., math, basketball, painting, etc.).

2) In a few sentences, describe how you learned this activity.

3) List two things you've done to become better at this activity.

4) Pair up with a student near you and describe your learning process for this activity (this should take around 1-2 minutes). When you are done, listen to your partners learning process and write down the two things they have done to become better at their activity.

Chem 111 – Week 3 In Class Assignment Name (STUDY SKILLS GROUP - B)

1) Write down one method you use when studying for this class.

2) In a few sentences, describe how you use this method.

3) List two other methods you either use or could use when studying for this class.

4) Pair up with a student near you and describe your study process (this should take around 1-2 minutes). When you are done, listen to your partners study process and write down two methods they use when studying.

APPENDIX E

COPYRIGHT PERMISSION FOR ARTICLE 1

Re: Permission to reprint manuscript

Prof Keith S Taber [kst24@cam.ac.uk]

Sent: Friday, March 25, 2016 3:13 PM

To: Ferrell, Brent

Cc: jbarbera@pdx.edu; cerp (shared) [cerp@rsc.org]

Dear Brent

I've checked out the current licence (which I do not think has changed recently) and I think the agreement you signed states:

The Author(s) may republish the Version of Record in theses of the Author(s) in printed form and may make available this PDF in the theses of Author(s) via any website(s) that the university(ies) of the Author(s) may have for the deposition of theses. No embargo period applies. Persons who receive or access the PDF mentioned above must be notified that this may not be further made available or distributed.

So I do not think you do need my permission (as long as you include a statement about not further copying or distributing), but if your graduate schools wants my approval, you have it.

I hope the examination/defence goes well.

Best wishes

Keith

On 24/03/2016 03:07, Ferrell, Brent wrote:

Dear Dr. Taber,

I am the lead author of the article "Analysis of students' self-efficacy, interest, and effort beliefs in general chemistry", DOI: 10.1039/C4RP00152D. I am requesting permission to reprint this manuscript as a chapter in my doctoral dissertation. Your permission is a requirement from the University of Northern Colorado graduate school in order to accept my dissertation. Please respond at your convenience.

Thank you,

Brent Ferrell, M.S.
Graduate Teaching Assistant
Department of Chemistry and Biochemistry
University of Northern Colorado
Ross Hall 3566

--