May 2017

Persistence in STEM: Development of a Persistence Model Integrating Self-efficacy, Outcome Expectations and Performance in Chemistry Gateway Courses

Shalini Srinivasan
University of Wisconsin-Milwaukee

PERSISTENCE IN STEM: DEVELOPMENT OF A PERSISTENCE MODEL INTEGRATING SELF-EFFICACY, OUTCOME EXPECTATIONS AND PERFORMANCE IN CHEMISTRY GATEWAY COURSES

by

Shalini Srinivasan

A Dissertation Submitted in

Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

in Chemistry

at

The University of Wisconsin – Milwaukee

May 2017

ABSTRACT

PERSISTENCE IN STEM: DEVELOPMENT OF A PERSISTENCE MODEL INTEGRATING SELF-EFFICACY, OUTCOME EXPECTATIONS AND PERFORMANCE IN CHEMISTRY GATEWAY COURSES

by

SHALINI SRINIVASAN

The University of Wisconsin-Milwaukee, 2017
Under the Supervision of Professor Kristen Murphy, PhD

STEM persistence has been an important issue, especially in the context of underrepresented groups based on race and gender. Researchers in the last decade or so have been examining the powerful impact that affective and cognitive factors can exert individually on performance and persistence. It is only reasonable to hypothesize that combining affective and cognitive measures would offer a more thorough understanding of factors that impact students' performance and STEM persistence. Evaluating these outcomes in the context of gateway courses is particularly essential due to the non-negligible percentage of students who drop out of these courses or decide to change their intended STEM majors after key testing events.

Using social cognitive career theory (SCCT) as a framework, this exploratory study set out to develop / adapt surveys to capture two key SCCT constructs – self-efficacy (SE) and outcome expectations (OE). These surveys were psychometrically tested and used in the development of cross-sectional predictive performance and persistence models for general chemistry. Items from both full-length surveys were subsequently used in the development of a shortened survey, which was administered as key points during a semester to evaluate changes in performance, SE or OE prior to or after testing events. Interventions, packaged as study tools, were also administered to

students before these events; the impact of these study tools on students' SE, OE and performance was also assessed in efforts to assemble preliminary profiles for at-risk students.

TABLE OF CONTENTS

PAGE

List of figures……..ix
List of tables……xi
List of abbreviations……xvi
Acknowledgements……xviii

CHAPTER

1 INTRODUCTION……...1
Purpose of the Study……3

2 LITERATURE REVIEW……4
Persistence in college……...4
Persistence and gender in STEM……5
Affect and gender in science education……...8
Social cognitive theory……..11
Self-efficacy……...12
Outcome expectations……15
Social Cognitive Career Theory (SCCT)……...16
SCCT model of interest development……17
SCCT model of career choice……18
SCCT model of task performance……..18
SCCT in STEM……..19
SCCT in studies with underrepresented populations……22

3 INSTRUMENTS AND METHODS……27
Norm-referenced instruments……27
Instruments and participants……..28
Mixed methods design……...28
Description of samples……...29
Preparatory chemistry……29
General Chemistry I and II (GC I and II)……..30
General chemistry for engineers……30
General scale development……30
Brief description of instruments……31
Methods and data analyses……34
Descriptive statistics……..34
Comparative statistics……34
Correlation analysis……...35
Factor analysis – Exploratory……36
Factor analysis – Confirmatory……40
Cluster analysis……..43
Psychometric theory……...44

Validity…………………………………………………………………..44
Reliability………………………………………………………………..45
Data cleaning……………………………………………………………46
Human subjects approval………………………………………………...46

4 DEVELOPMENT AND VALIDATION OF THE CHEMISTRY SELF-EFFICACY AND ANXIETY SURVEY (CSEAS)…………………………………………………...48
Background and rationale………………………………………………………..48
Methodology……………………………………………………………………..52
Development of the CSEAS items – self-efficacy and anxiety…………52
Structure of the CSEAS…………………………………………………55
Student interviews – instrument development and implementation……..57
Data collection and participants…………………………………………61
Data analyses……………………………………………………………………64
Results and discussion…………………………………………………………...68
Semantic differential……………………………………………………..68
Descriptive statistics……………………………………………………..68
Factor analysis – Exploratory……………………………………………69
Factor analysis – Confirmatory…………………………………………77
Comparative statistics……………………………………………………81
Reliability………………………………………………………………...83
Validity…………………………………………………………………..84
Cluster analyses…………………………………………………………94
Limitations………………………………………………………………………97
Conclusions and implications……………………………………………………98

5 DEVELOPMENT AND VALIDATION OF THE CHEMISTRY OUTCOME EXPECTATIONS SURVEY (COES)…………………………………………………102
Background and rationale………………………………………………………102
Methodology……………………………………………………………………104
Development of the COES items………………………………………104
Structure of the COES…………………………………………………108
Student interviews – instrument development and implementation……109
Data collection and participants………………………………………...110
Data analyses…………………………………………………………………...112
Results and discussion…………………………………………………………113
Descriptive statistics……………………………………………………113
Factor analysis – Exploratory…………………………………………..115
Factor analysis – Confirmatory…………………………………………123
Comparative statistics…………………………………………………..125
Reliability………………………………………………………………129
Validity…………………………………………………………………131
Cluster analyses………………………………………………………...140
Limitations……………………………………………………………………...145
Conclusions and implications…………………………………………………..146

6 DEVELOPMENT AND VALIDATION OF A SHORTENED INSTRUMENT TO MEASURE SELF-EFFICACY OUTCOME EXPECTATIONS……………………149
Background and rationale……………………149
Methodology……………………151
Item selection for the subset survey……………………151
Structure of the subset instrument……………………154
Student interviews……………………157
Data collection and participants……………………158
Data analyses……………………161
Comparative statistics……………………161
Psychometric testing – construct validity……………………162
Predictive validity – standard multiple linear regression (SMLR)……………………162
Predicting group membership – discriminant analysis……………………165
Results and discussion……………………167
Psychometric testing – exploratory factor analysis (EFA)……………………173
Standard multiple linear regression (SMLR)……………………176
Discriminant analysis (DA)……………………182
Limitations……………………190
Conclusions and implications……………………192

7 DESIGN AND TESTING OF INTERVENTIONS IN GENERAL CHEMISTRY……………………194
Background and rationale……………………194
Methodology……………………196
Design of the intervention module……………………196
Testing interventions – classwide usability study……………………201
Data collection and participants……………………205
Data analyses……………………206
Predictive validity – standard multiple linear regression (SMLR)……………………206
Predicting group membership – discriminant analysis……………………206
Odds ratios (ORs)……………………207
Results and discussion……………………208
Standard multiple linear regression (SMLR)……………………209
Predicting group membership – discriminant analysis……………………211
Odds ratios……………………212
Usability studies – eye tracking and student interviews……………………214
Limitations……………………216
Conclusions and implications……………………217

8 DEVELOPMENT OF PERFORMANCE AND STEM PERSISTENCE MODELS IN GENERAL CHEMISTRY GATEWAY COURSES……………………219
Introduction……………………219
Purpose of the study……………………220
Research methodology……………………221
Research design……………………221
Participants……………………224

Data analysis – performance model……………………………………….226
Data analysis – persistence model……………………………………...227
Results and discussion………………………………………………………….228
Cross-sectional performance model – GC I final exam as outcome……228
Cross-sectional performance model – GC I course performance as outcome………………………………………………………………235
Cross-sectional persistence model……………………………………...239
Limitations……………………………………………………………………...246
Conclusions and implications…………………………………………………..247

9 OVERALL CONCLUSIONS AND IMPLICATIONS………………………………...250
Conclusions……………………………………………………………………..250
Instrument development and psychometric testing……………………..251
Model testing at key time points………………………………………..252
Methodological limitations……………………………………………………..253
Implications……………………………………………………………………..255
Implications for teaching……………………………………………….255
Implications for chemical education research…………………………..256
Future research………………………………………………………………….257

REFERENCES…………………………………………………………………………...259

APPENDICES…………………………………………………………………………….274
Appendix A: SAS code used for confirmatory factor analysis (CFA)…………274
Appendix B: IRB consent form – class-wide data collection…………………..275
Appendix C: IRB consent form – think aloud/interviews……………………...276
Appendix D: Fall 2012 self-efficacy standalone survey………………………..278
Appendix E: Fall 2012 – survey version with stress and self-efficacy scales….279
Appendix F: Fall 2012 - Anxiety survey (standalone).....……………………….281
Appendix G: S13 and after – CSEAS online version (Qualtrics)………………282
Appendix H: Results summary – semantic differential (Spring 2012, GC I)…..287
Appendix I: CSEQ survey……………………………………………………...295
Appendix J: Pre-COES (outcome expectations) scale………………………….296
Appendix K: Post-COES (outcome expectations) scale………………………..298
Appendix L: Subset (shortened) instrument……………………………………300
Appendix M: STEM major designations – as of May 2016……………………302
Appendix N: Results of one-way ANOVA to select logistic regression predictors……………………………………………………...306

Curriculum Vitae………………………………………………………………307

LIST OF FIGURES

Figure 2.1. Social Cognitive Career Theory Model……………………………………………..17

Figure 2.2. SCCT Model of task performance…………………………………………………..19

Figure 4.1. Example of notecards used for sorting task during interviews……………………...60

Figure 4.2. Example of notecards sorted into meaningful groups……………………………….61

Figure 4.3. Example of notecards with actual survey items and numbers for sorting task during interviews……………………………………………………………………………………61

Figure 4.4. Plot of frequency of students against online survey completion time for Fall 2014……………………………………………………………………………………66

Figure 4.5. Scree plot showing eigenvalues for retaining a four- or five-factor solution ………74

Figure 4.6. Standardized coefficients for the final, refined six-factor model of the chemistry self-efficacy and anxiety scale……………………………………………………………80

Figure 5.1. Scree plot showing eigenvalues for a five-factor solution…………………………119

Figure 5.2. Standardized coefficients for the final, refined five-factor model of the chemistry outcome expectations scale………………………………………………………125

Figure 6.1. Subset administration (GC I and GC II) showing time points and potential triggers...159

Figure 6.2. Cursory trends in outcome expectations (subset survey) across semesters (AY14-AY15) in GC I…………………………………………………………………..171

Figure 6.3. Cursory trends in outcome expectations (subset survey) across semesters (AY14-AY15) in GC I……………………………………………………………………172

Figure 6.4. Scatter plot distribution of residuals for a performance model using scores on Exam 1 – GCI…………………………………………………………………………...180

Figure 8.1. Outcome and predictors for developing and testing the performance model……...222

Figure 8.2. Outcome and predictors for developing and testing the persistence model……….223

Figure 8.3. Scatter plot distribution of residuals for performance model using GC I final exam Scores…………………………………………………………………………....234

Figure 8.4. Scatter plot distribution of residuals for performance model using GC I course percentages……...238

LIST OF TABLES

Table 2.1. Top supports and barriers in Math for male and female students at three educational levels……………………………………………………………………………...23

Table 2.2. Top supports and barriers in Science for male and female students at three educational levels……………………………………………………………………24

Table 3.1. List of instruments and administration details………………………………………..28

Table 4.1. Summary of testing purpose, timeline and data collected - CSEAS…………………62

Table 4.2. Participants (by course) for pilot administration of paper version of CSEAS – Fall 2012……………………………………………………………………………..64

Table 4.3. Participants (by course) for administration of online version of CSEAS – Spring 2013……………………………………………………………………………..64

Table 4.4. Demographic characteristics of a) combined datasets from Fall 2012, Spring 2013 and post- GC I from Fall 2014……………………………………………………68

Table 4.5. Component matrices: Factor structures of datasets from Fall 2012, Spring 2013 and Fall 2014………………………………………………………………………..70

Table 4.6. Correlation matrix, means and standard deviations for the chemistry self-efficacy component of the CSEAS…………………………………………………………73

Table 4.7. Parallel analysis results showing eigenvalues for actual and random ordered data – CSEAS…………………………………………………………………...………..74

Table 4.8. Rotated component matrix for four-factor solution: CSEAS……………….…….76

Table 4.9. Component matrix: Four-factor solution for the anxiety scale from CSEAS………..77

Table 4.10. Goodness-of-fit indicators of models for CSEAS at three time points during AY15-16………………………………………………………………………….…………78

Table 4.11. CSEAS subscale scores showing pre to post changes for GC I…………..………...81

Table 4.12. Results showing differences in CSEAS subscales based on gender………………...82

Table 4.13. Factor correlations and Cronbach alpha coefficients for the CSEAS…………..…..84

Table 4.14. Correlations between pre-CSEAS subscale scores and performance indicators in GC I……………………………………………………………………….…………85

Table 4.15. Correlations between post- CSEAS subscale scores and performance indicators in GC I……………………………………………………………………………….86

Table 4.16. Results showing differences in subscales based on high vs. low performing groups on the final exam – CSEAS……………………………………………..…….87

Table 4.17. Results showing gender differences in subscales based on low performance on the final exam-CSEAS……………………………………………………………88

Table 4.18. Results showing differences in subscales based on high vs. low performing groups in the course-CSEAS…………………………………………………………88

Table 4.19. Results showing gender differences in subscales based on low performance in the course-CSEAS……………………………………………………….………..89

Table 4.20. Correlations between self-efficacy and anxiety subscales in the CSEAS – GC I…..90

Table 4.21. Example of student generated CSEAS item groupings – GC I………………………92

Table 4.22. Three-cluster solution for preparatory chemistry…………………………………..95

Table 4.23. Three-cluster solution for general chemistry for engineers…………………..…….96

Table 5.1. Examples of revisions made to items from ChemX for COES item development………………………………………………………………..…….107

Table 5.2. Participants (by course) for pilot administration of COES – Fall 2013…………….111

Table 5.3. Participants (by course) for main administration of COES – Spring 2014………...112

Table 5.4. Demographic characteristics of a) post-GC I, b) pre-GC II and c) the combined dataset from Spring 2014……………………………………………...………113

Table 5.5. Summary of datasets - number of students, course and time point for analyses……...115

Table 5.6. Component matrix: Six-factor solutions for post GC I and pre GC II……………..116

Table 5.7. Correlation matrix, means and standard deviations for the chemistry outcome expectations scale (COES)…………………………………………………………118

Table 5.8. Parallel analysis results showing eigenvalues for actual and random ordered data – COES…………………………………………………………………..………120

Table 5.9. Rotated component matrix for five-factor solution: COES………………..……….122

Table 5.10. Goodness-of-fit indicators for COES models tested at three time points during AY14-15……………………………………………………………………….123

Table 5.11. COES scores showing pre to post changes for GC I…………………...………126

Table 5.12. Results showing differences in COES pre-subscales based on gender…………..127

Table 5.13. Results showing gender differences within subscales for low ability students…...128

Table 5.14. Factor correlations and Cronbach alpha coefficients for the COES……………….129

Table 5.15. Correlations between pre- COES subscale scores and performance indicators in GC I……………………………………………………………………...……….131

Table 5.16. Correlations between post- COES subscale scores and performance indicators in GC I……………………………………………………………………….………..133

Table 5.17. Results showing differences in COES pre-subscales based on high vs. low performing groups on the final exam…………………………………...……….133

Table 5.18. Results showing differences in COES pre-subscales based on high vs. low performing groups in the course……………………………………….……….134

Table 5.19. Correlations between pre-COES and pre-CSEAS subscale scores in GC I……………………………………………………………………...……….136

Table 5.20. Correlations between post-COES and post-CSEAS subscale scores in GC I…….137

Table 5.21. Four-cluster solution for preparatory chemistry students – COES………………..143

Table 5.22. Four-cluster solution for general chemistry for engineers – COES……………….144

Table 6.1. Subscales resulting from the CSEAS……………………………………………….152

Table 6.2. Subscales resulting from the COES. ………………………………………………..152

Table 6.3. COES subscales and their associated subset items…………………………………..155

Table 6.4. CSEAS subscales and their associated subset items………………………………..156

Table 6.5. GC I participants (by gender) for different time points of subset administration……..161

Table 6.6. Predictors and outcome variable for each subset performance model (GC I)………..163

Table 6.7. Descriptive statistics for self-efficacy subscales at each time point during subset survey administration……………………………………………………………...………..168

Table 6.8. Descriptive statistics for outcome expectations subscales at each time point during subset survey administration……………………………………………………..169

Table 6.9. Self-efficacy factor structures (tracking item changes) from subset survey administration across a semester (AY14-15) in GC I…………………...………174

Table 6.10. Outcome expectations factor structures (tracking item changes) from subset survey administration across a semester (AY14-15) in GC I………………………..…….175

Table 6.11. Bivariate correlations between predictor variables and score on Exam 1 for GCI (AY14-15)……………………………………………………………………….……177

Table 6.12. Summary of multiple regression analysis for students' Exam 1 scores in GC I…..179

Table 6.13. Summary of multiple regression analysis for students' Exam 2 scores in GC I…..181

Table 6.14. Summary of multiple regression analysis for students' Exam 3 scores in GC I…..182

Table 6.15. Point-biserial correlations between predictor variables and performance group on exam 1 for GCI (AY14-15)……………………………………………………………..184

Table 6.16. Tests of equality of group means - DA model predicting group membership in exam 1 – GCI……………………………………………………………...……….185

Table 6.17. Box's test of equality of covariance matrices………………………………………186

Table 6.18. Summary of canonical discriminant functions for membership in performance groups – GC I Exam 1…………………………………………………………………..….186

Table 6.19. Unstandardized canonical discriminant function coefficients…………………….187

Table 6.20. Classification statistics showing correct and incorrect classification of cases – Exam I……………………………………………………………………188

Table 7.1. Example of problem solving strategies contained in intervention module – GC I….199

Table 7.2. GC I intervention participants (by gender) for three testing events………………...205

Table 7.3. Example of set up for calculating odds ratios (OR)………………………………...207

Table 7.4. Descriptive statistics for outcome expectations subscales before and after testing events – Intervention group in GC I (S16-F16)…………………………………….208

Table 7.5. Descriptive statistics for self-efficacy subscales before and after testing events – Intervention group in GC I (S16-F16)……………………………...…………209

Table 7.6. Summary of predictive model residuals (control and treatment datasets) and regression equations – GCI……………………………………………………….……….210

Table 7.7. Combination of predictor variables and resulting predictive model residuals - GC I (F16)…………………………………………….………..210

Table 7.8. Classification table for treatment (intervention) dataset – GC I (S16+F16)…………211

Table 8.1. Persistence categories and codes utilized for logistic regression…………………..224

Table 8.2. Descriptive statistics (by gender) for variables in the performance model (data from S14 – F15)…………………………………………………………….229

Table 8.3. Bivariate correlations between predictor variables and GC I final exam % for cross-sectional performance model……………………………………………………..231

Table 8.4. Summary of multiple regression analysis for students' final exam scores in GC I……………………………………………………………………………233

Table 8.5. Bivariate correlations between predictor variables and GC I course % for cross-sectional performance model…………………………………………….………235

Table 8.6. Summary of multiple regression analysis for students' course performance in GC I……………………………………………………………………………237

Table 8.7a. Descriptive statistics (by category) for variables in the persistence model……….239

Table 8.7b (Continued). Descriptive statistics (by category) for variables in the persistence model…………………………………………………………………………..240

Table 8.8. Crosstab analysis showing population of each category in the GC I persistence model…………………………………………………………………………..241

Table 8.9. Crosstab analysis showing population of each category (by gender) in the GC I persistence model…………………………………………………….…..241

Table 8.10. Model fitting information for GC I cross sectional persistence model……………...242

Table 8.11. Likelihood ratio tests to indicate importance of predictors in the persistence model……………………………………………………………..…243

Table 8.12. Parameter estimates for predictors in the GC I persistence model………..………..244

Table 8.13. Classification table showing utility of model in categorizing cases………..………245

LIST OF ABBREVIATIONS

STEM	Science, technology, engineering and mathematics
pSTEM	Physical science, technology, engineering and mathematics
SCT	Social cognitive theory
SCCT	Social cognitive career theory
COES	Chemistry outcome expectations survey
CSEAS	Chemistry self-efficacy and anxiety survey
OE	Outcome expectations
SE	Self-efficacy
TP/TCPE	Toledo placement / Toledo chemistry placement exam
GC	General chemistry
ACS	American Chemical Society
ACT	American College Testing
GPA	Grade point average
EFA	Exploratory factor analysis
CFA	Confirmatory factor analysis
PCA	Principal component analysis
PAF	Principal axis factoring
ML	Maximum likelihood
FIML	Full information maximum likelihood
LM	Lagrange multiplier
SRMR	Standardized root mean square residual
TLI	Tucker-Lewis index

RMSEA	Root mean square error of approximation
IRB	Institutional review board
CAEQ	Chemistry attitudes and experiences questionnaire
MSES	Mathematics self-efficacy scale
CSSS	Chemistry self-efficacy scale for college students
PA	Parallel analysis
ITC	Item total correlation
CHEMX	Chemistry Expectations survey
CSEQ	College Student Experiences Questionnaire
KMO statistic	Kaiser-Meyer-Olkin statistic
SMLR	Standard multiple linear regression
DA	Discriminant analysis
AOI	Area of interest
OR	Odds ratio
VIF	Variance inflation factor
ANOVA	Analysis of variance

ACKNOWLEDGEMENTS

This process was by no means a solo endeavor. Thanks are due in small and large doses to several individuals who facilitated this quest. I'd like to start by thanking my advisor Dr. Kristen Murphy. The scope of this work, the sheer ambition behind it and its inception are the products of her brilliant mind. A heartfelt thank you for being a true mentor! I'm always in awe and will continue to be in awe of the wizardry that is your multitasking (*and then some*) life. Thanks for the patient guidance and encouragement throughout this process.

Thanks to my committee – Dr. Anja Blecking, Dr. Scott Lewis at the University of South Florida, Dr. Andy Pacheco and Dr. Jorge Woehl – for being patient and flexible in scheduling milestone meetings, orals and my defense. Thanks for your consideration, comments and insight into my project.

Thanks to Dr. Jamie Schneider at UW-River Falls for the opportunity to be a part of your research project each summer and during my last semester at UWM.

I'd also like to thank my research group (past and recent members) – Dr. Lisa Kendhammer, soon to be Dr. Jaclyn Trate, Karri Gerlach, Erin O'Connell and Victoria Fisher – for all your help and support at various stages of this process. Thanks to all the undergraduate student employees for transcribing interviews and contributing to aspects of this project.

Much thanks to everyone in the chemistry department at UWM – faculty, staff, fellow graduate students and undergraduate – for making this journey enriching, sometimes frustrating (I'm looking at you, lab report grading and 105), but enjoyable nevertheless. Special thanks to instructors – Dr. Christine Carlson, Dr. Thomas Sorensen and Dr. Anja Blecking – for allowing me to collect research data in your courses. This dissertation would have a few pages of missing analyses had it not been for your efforts. Many thanks to the current and former office staff –

Elise, Mary, Wendy, Shelly, Megan, Mike, Annette and Kevin – for assisting with administrative matters, large or small.

To my family – the gratitude is ongoing and implied, but thanks for everything. This would not have happened without your relentless support and prayers. Thanks to my mom for being a worrier so I'd worry less and thanks to the eternal optimists in my life – my dad, sister and Radha (affectionately called Aunt Kay). Thanks to all the uncles and aunts in the family for your support. I'll need a separate chapter to describe how fantastic all of you are and how much I love all of you. Thanks for being a part of this journey.

From a personal standpoint, this journey would have been impossible without the largesse of the following people: Hari Banda and Gauri Srinivasan. Thank you for your invaluable help and support. UWM would have not been possible for me if it weren't for your generosity.

A huge thanks to old friends and new, especially a certain Wauwatosa resident, for fun and memorable discussions about life, family, Tolkien, the shared disdain for hot weather, Asylum movies (*complete with subplots*), Game of Thrones (*Bran is why we can't have nice things!*) and of course, Archer. Thank you Sandra, *other Sandra*, Deddy and Mello.

As an animal lover, I can't help but thank a little *furball* who would barge into my room to wake me up, wait for me by the door and greet me ever so affectionately when I got home from work. Many thanks and love to Symba for being the ray of sunshine during the somewhat stressful writing process and my free shrink at times. Thanks to Karin for being you!

Lastly, thanks to my former husband. Closing this door - literally and figuratively - opened the door to a doctoral program at UWM. So, thank you.

CHAPTER 1: INTRODUCTION

Over the past several decades, there has been a substantial decrease in the percentage of science, technology, engineering and mathematics (STEM) majors relative to the overall undergraduate population (PCAST, 2012). One of the major sources of this decline has been students' lack of persistence in their intended STEM majors; about 60% of students who enroll in a STEM field switch to a non-STEM field or drop out of their degree program entirely (PCAST, 2012; Waldrop, 2015). The numbers become even more alarming when underrepresented student groups – females, racial and ethnic minorities – are considered. About a third of the catalogued federal funding for STEM education is geared towards increasing the participation of underrepresented groups in STEM careers, with about 10% of that funding explicitly directed towards females in STEM education (PCAST, 2012). While the gender gap has narrowed in the physical sciences, in engineering, half the female students leave the field while only 10% of the male students leave (Singh et al., 2013). Thus, much of the research in STEM persistence and any effective interventions have focused on initiatives to help introduce female students to math and engineering as careers. Given that the United States workforce will face a deficit of one million college graduates in STEM over the next decade or so, it is imperative to address the very real problem of persistence in STEM and develop a model to help identify factors that contribute to a lack of persistence (PCAST, 2012).

Taking up this phenomenon of STEM persistence in the chemistry domain is essential because similar to engineering and math, a vast majority of students intending to major in STEM fields enroll in chemistry gateway courses during the first two years of their program; these years mark a critical decision point to "switch" or persist in STEM majors. The first year is especially important because 35 % of STEM majors "switch" after their first year (Business-Higher Education Forum, 2010); besides, a non-negligible percentage of students attempting introductory chemistry drop the course or change to a non-STEM major (PCAST, 2012). While the reasons for understanding these phenomena have almost always focused on the cognitive domain, investigating STEM persistence has necessitated the exploration of domains beyond the academic. Regardless of ability, students' interests, motivations and beliefs about themselves have a fairly strong impact on whether they "persist in", "switch out of" or leave STEM fields entirely (Seymour & Hewitt, 1997). Discipline-based education researchers interested in retention and representation issues have for some time shifted their focus to the affective domain to better understand persistence and the participation differential in STEM.

Social cognitive career theory (SCCT) has become a frequently used framework for studying academic and career development. The performance model in SCCT lists five distinct, yet bi-directionally related cognitive and affective variables that influence academic performance and persistence: Past performance, ability, outcome expectations, self-efficacy beliefs and goal mechanisms (Brown et al., 2008). Two key constructs (factors) to emerge from this framework were self-efficacy and outcome expectations. While individual instruments to measure self-efficacy have been developed and tested in most science disciplines, including chemistry, outcome expectation measures have received minimal attention. Ideally, the measures for persistence should be merged with performance measures and tested on an entire sample of students and key subgroups within that sample. However, the scarcity of psychometrically viable outcome expectations measures has limited the development and testing of comprehensive, longitudinal performance and persistence models. In addition, while these models have been tested empirically, their predictive utility has not been examined on a finer grain to (a) identify students at risk for lack of persistence, (b) identify the point at which there is a decline in persistence or performance measures and (c) implement an appropriate intervention to target students and offset their lack of persistence.

Purpose of the Study

The objective of this study was to investigate the impact of performance, self-efficacy and outcome expectations on persistence of students in STEM majors during their enrollment in general chemistry gateway courses. As part of this objective:

1) This study aimed to develop a valid and reliable instrument that could be used to measure chemistry outcome expectations (COES) in first-year chemistry courses.
2) A valid and reliable chemistry self-efficacy instrument (CSEAS) was adapted specifically for this study.
3) Models of performance (content based and course performance) were tested locally within a course to identify predictors that would influence chemistry performance.
4) Affective measures, in combination with performance, were used in the development of persistence models which categorized membership of students based on whether they persisted in their intended STEM majors while enrolled in a course.
5) Based on changes in the pre-post affective measures, a subset instrument was developed to capture changes in performance and affective measures on a much finer grain; this also allowed for identification of triggers and points at which a performance or affective measured dropped.
6) Finally, based on the profiles and changes indicated by the subset instrument, interventions were developed, utilized and tested in an effort to offset the decline in measures of performance, persistence or both.

CHAPTER 2: LITERATURE REVIEW

This chapter reviews the literature and theories of persistence, both in college and STEM in particular, that have shaped this research. Seminal works that have addressed the historic underrepresentation of females in science will also be examined. In addition, this review offers an insight into the cognitive and affective domains and how these domains and resulting theories have not only helped understand and explain persistence and the participation differential in STEM, but also provided a framework for this study.

Persistence in College

For several decades now, student retention has been an extremely important goal for higher education institutions and scholars alike. While studies examining '*retention*' and '*persistence*' have been abundant, these terms have often been used ambiguously and in some cases mistakenly interchangeably (Reason, 2009). Retention is an "organizational phenomenon" used to described the idea of educational institutions retaining students. On the contrary, persistence is an "individual phenomenon", which describes students' intentions to "persist to a goal" (Reason, 2009). These goals could be proximal (completing a course) or distal (completing a degree), offering an added distinction between the phenomena of retention and persistence (Reason, 2009).

Two of the earliest and fairly important models to explain college persistence were Vincent Tinto's Student Integration Model (SIM) (Tinto, 1975) and Bean's Student Attrition Model (Bean, 1980). While Tinto's model is based on the extent to which a student is socially and academically integrated into an institution, the Student Attrition Model examines the role of intentions, attitudes and external factors as predictors of persistence. Despite the scarcity in studies that tested the predictive validity of these models, several subsequent studies utilized these models as theoretical frameworks for studying aspects of college persistence (Cabrera, Castañeda, Nora & Hengstler, 1992). While retention rates certainly improved during the 1990s, the collective impact on graduation rates remained minimal and although these seminal models provided a basis for persistence studies, much of the work does not account for the interrelated influences that dictate student persistence.

More recently, Reason and Terenzini compiled a comprehensive review of persistence research and developed a conceptual framework that integrated students' precollege characteristics, their peer environment, institutional characteristics and individual student experience (Terenzini & Reason, 2005; Reason, 2009). While this framework definitely provided a more comprehensive examination of college persistence in general, viewing this phenomenon through sociodemographic and psychosocial lenses was offering some alarming results of its own. Degrees in STEM (science, technology, engineering and mathematics) fields were not being awarded at the same rate as the total numbers of bachelor's degrees in the United States; in 2006, the percentage of students graduating with degrees in STEM was no different (or lower) than that in previous years (National Science Foundation [NSF], 2010; Maltese & Tai, 2011). Moreover, there was a crucial need to address norms and practices which made it difficult for underrepresented minorities, especially female students, to persist (Seymour & Hewitt, 1997).

Persistence and gender in STEM

Concerns about the nature of science and mathematics education started coming into the spotlight in the mid-1980s when the Higher Education Research Institute brought attention to the waning percentage of freshmen opting to enter and persist in science and mathematics-based majors (Dey, Astin & Korn, 1991; Astin & Astin, 1993; Astin et. al., 1985; Seymour & Hewitt, 1997). Much of the debate surrounding these issues was motivated by knowledge that the general population lacked science literacy due to the disappointing efforts made in science and mathematics education; moreover, not enough was being done to recruit and retain students in these fields and any recruitment was severely biased in terms of gender and race (Seymour & Hewitt, 1997).

To clarify terminology at this point, the terms gender and sex have been used interchangeably and incorrectly in most literature on social sciences, including affective research. The term sex refers to a biological construct and defines an individual as male or female based on genetics, anatomy and physiology (Tannenbaum et al., 2016). On the other hand, gender refers to a multifaceted and fluid construct, impacted by social and cultural contexts and environments (Tannenbaum et al., 2016). As gender is a fairly broad term that can also indicate the identities of girls, women, boys and men, the definitions of sex and gender are changing and are often interrelated (Tannenbaum et al., 2016). However, for the implementation of research methods or reporting outcomes by males vs. females, the term sex is deemed more appropriate. While this dissertation uses the delineations of males and females, these categories appear under the term 'gender' in demographic data sought through institutional research and will be used as part of this term (as opposed to the correct terminology 'sex') to stay consistent.

Factors that have shown to contribute to the persistence of females and minorities in STEM fields range from institutional policies, preparation in high school and college to financial assistance (May & Chubin, 2003; George-Jackson, 2011). Students' scores on standardized tests and their performance in high school math and science courses have been known to predict college-level math and science performance in addition to persistence in STEM fields (Elliot et al., 1996; George-Jackson, 2011). Differential decisions made by male and female students about their majors have also been attributed to academic performance wherein male and female students respond differently to failing a course in their major. A study conducted using chemical engineering students revealed that male students were more likely to retake a failing course while female students were likely to seek out a new major (Felder et al., 1995).

On the heels of these statistics, the President's Council of Advisors on Science and Technology released a report in 2012 in which the problem of persistence and preparedness in STEM subjects was addressed. The report stated that if the United States had to sustain its position as leader in research and development, it must produce approximately one million more workers in STEM fields over the next decade (PCAST, 2012). While this report reinforced the idea of a general lack of persistence in STEM fields, it emphasized the serious underrepresentation of females in STEM and the need for opportunities that would encourage and allow females to fully participate in exciting STEM experiences.

The issue of underrepresentation of females in STEM has been prevalent since the 1970s when the metaphor "leaky pipeline" was used to describe the relatively high attrition of females from STEM fields at multiple time points during their academic tenures (Miller & Wai, 2015; Berryman, 1983; Alper, 1993). Studies have shown that this metaphor, while useful at the time, has in fact revealed inconsistencies due to the changing landscape in the 1990s when the gender gap narrowed among STEM bachelor's degree earners. Other studies have shown that the persistence differential exists only in some STEM fields. Recently, a 30-year retrospective analysis investigating empirical support for the "leaky pipeline" revealed its utility in partially explaining historical gender differences, but suggested that the metaphor is not very applicable to current gender differences in the transition from STEM bachelors to Ph.D. programs as persistence rates have converged in several STEM fields. (Miller & Wai, 2015). In the 1970s, male students earning pSTEM bachelor's degrees were 1.6 to 1.7 times as likely as females to later earn a pSTEM Ph.D; however, this gap completely closed by the 1990s (Miller & Wai, 2015). In general, the utilization of this metaphor ignores factors such as a student's entry into STEM before pursuing a bachelor's degree or that successful completion of a STEM degree occurs even if a student has not navigated the traditional STEM "pipeline" (Miller & Wai, 2015).

Regardless of the studies that suggest convergence in persistence rates between males and females, the literature on gender differences in STEM fields is conflicting and requires further investigation. Furthermore, while the factors that impact a student's pursuit of, persistence in and departure from STEM fields have almost always involved the cognitive domain, several researchers over the last decade or so have been looking at non-cognitive factors to understand academic performance and persistence. Moreover, these non-cognitive factors manifest themselves in various ways, resulting in performance or persistence differentials among student subgroups.

Affect and gender in science education

The origins of the domains of learning can be traced to the period between 1956 – 1972 when a group of educators made unique contributions to the development and refinement of each domain. Benjamin Bloom started examining educational objectives by exploring the cognitive domain, which has been the main focus of curricula and involves the development of intellectual skills (Bloom et. al, 1956). Krathwohl's taxonomy focused on the affective domain, which examines emotional and behavioral outcomes such as feelings, motivations and attitudes (Krathwohl, Bloom & Masia, 1973). Various versions of taxonomies pertaining to the psychomotor domain, which involves development of motor-skills and coordination, were developed by Simpson, Harrow and Dave (Dave, 1970; Harrow, 1972 & Simpson, 1972). While research integrating all three domains is limited, studies have examined each domain or combinations quite comprehensively in science education overall, and within specific divisions in science education.

Work in the affective domain dates back to the early 1960s when a comprehensive inquiry was initiated to evaluate the number of students entering science and technology in higher education. The phenomenon, known as the 'swing from science' was attributed to declining interest in science and general dissatisfaction among science students (Dainton, 1968; Osborne, Simon & Collins, 2003). This led to a plethora of work over the past forty years by the science education research community, with much of the work heavily focused on students' attitudes towards science (Osborne, Simon & Collins, 2003). Over the years, the definition of 'attitudes' has been amended to include several sub-constructs - such as anxiety, self-esteem and motivation - that proportionally contribute towards an individual's attitude towards science (Osborne, Simon & Collins, 2003). Attitude toward science has been shown to influence achievement, choice of science courses and careers (Napier & Riley, 1985; Germann, 1988).

Investigation of the factors that impact students' attitudes towards science revealed that gender was a key contributor towards students' attitudes; much of the research conducted in the early 1990s showed that boys had a more positive attitude to science than girls. A report published by the National Education Goals (1993) stated that positive attitudes toward science and mathematics were more likely to be demonstrated by students in higher grade levels, with large gaps between male and female students. Although this trend changed in the late 1990s and gender did not play a major role in achieving success, female students with high abilities and confidence were still opting out of pursuing science fields due to the uninspiring nature of these fields (Osborne, Simon & Collins, 2003). This effect was further emphasized when Seymour and Hewitt conducted their ethnographic project to examine science, mathematics and engineering (SME) students' reflections on their undergraduate experiences and determine the reasons for attrition and persistence; while students who switched out of ('*switchers*') or stayed in ('*non-switchers*') STEM displayed the same range of abilities, motivations and behaviors, those who persisted shared distinct attitudes and coping skills including confidence, a strong, sustained interest in their intended fields of study and in their careers and a strong support system, especially for female students. Reasons cited by students who left STEM fields included poor quality of teaching, a "chilly classroom climate", lack of faculty-student interaction, lack of preparation and discouragement at academic challenges (Seymour & Hewitt, 1997; Hall & Sandler, 1982). While Seymour and Hewitt focused primarily on students who have already entered STEM fields, their study signaled a shift in emerging research in which educators started to address factors impacting STEM career choices and underrepresentation at the high school level as well (Williams & Ceci, 2007; Hughes, 2011).

Several researchers investigated the effects of "rigorous" high school coursework and concluded that the rigor of a high school program had a significant impact on students' attainment of college degrees (Adelman, 2006; Horn & Kojaku, 2001; Trusty, 2002; Tyson et al. 2007; Maltese & Tai, 2010). High school grade point average (GPA) and high educational aspirations were also positively associated with male and female students majoring in STEM (Ware & Lee, 1988). Mau (2003) examined the effects of gender and race on the stability of aspirations to follow a career in science and engineering; the conclusions revealed that gender, race, mathematics self-efficacy and academic proficiency were key factors in the persistence of career aspirations. What these studies emphasized was the need to examine factors beyond performance (cognition) to provide a better understanding of and model for learning, performance, vocational choices and persistence. The social context, interplay between self-beliefs and environment, self-regulation and the idea that learning takes place even in the absence of an observable response have become important concepts critical for learning and modeling both performance and persistence (Bandura, 1986).

Social Cognitive Theory

In 1986, Albert Bandura developed social cognitive theory (SCT) as a way to explain human behavior. Deviating from traditional behaviorist theories – in which situational and cognitive influences are mostly ignored (Bandura, 1977) – the SCT framework sought to explain behavior as a mechanism in which external environmental factors, overt behavior and personal agency (in the form of cognition, affect and biological events) function as interacting components that reciprocally impact each other as well. This model – known as “triadic reciprocal causality” – forms the basis of SCT (Bandura, 1986). The implication is that while individuals can exercise personal agency, they are constrained by external consequences, their own experiences and self-reflective processes. Self-referent thought mediates knowledge and action and the strength of self-regulatory processes determines what courses of action are pursued (Bandura, 1986). The ways in which individuals interpret their goals or attainments impact how their environments and self-beliefs might be altered; these interpretations subsequently alter their future performance goals (Pajares, 1996). Among the various personal determinants of psychosocial functioning, three mechanisms have been influential in understanding vocational choices and career development. The concepts of self-efficacy, outcome expectations and goals and the relationships among these concepts form the core of the theoretical framework that shapes the research discussed in this dissertation.

Self-efficacy

Self-efficacy refers to “people’s judgments of their capabilities to organize and execute courses of action required to attain designated types of performance” (Bandura, 1986, p. 391); quite simply, it answers the question “can I do this?”. Bandura theorized four sources of self-efficacy: Performance accomplishments, vicarious learning or modeling, verbal persuasion and emotional arousal in relation to the behavior. Performance outcomes, described as an individual’s successes and failures, and past experiences are expected to have the greatest influence on self-efficacy beliefs. Success in a task increases a person’s confidence to perform another similar task while failure correspondingly decreases their self-efficacy (Bandura, 1986). However, if individuals can see these failures as attainable challenges and overcome them by conviction, they can increase their self-motivated persistence (Bandura, 1977). Individuals also form self-efficacy beliefs through vicarious experiences such as watching peers succeed or fail. According to Bandura, vicarious experiences can have a larger impact on a person’s self-efficacy if the person has less experience in a task and consequently less stability in their self-efficacy beliefs (Bandura, 1986). These conditions can also be helped by verbal encouragement or verbal persuasions, the third source of self-efficacy. The expertise and credibility of the ‘persuader’ are factors that impact the effectiveness of persuasions. Thus, if an individual attempts a task due to verbal persuasion and fails, the ‘persuader’ may be discredited (Bandura, 1986). Lastly, emotional states such as anxiety, stress and fatigue also impact self-efficacy beliefs. Judgments of self-efficacy are not directly predicated by these sources; instead they are highly dependent on the manner in which a person combines these sources to select, integrate, interpret and recollect information. Consequently, making judgments of self-efficacy is a highly personal and person specific process (Bandura, 1986).

Often considered the most pervasive factor of personal agency, self-efficacy beliefs have received increasing attention in educational research, especially in studies involving academic motivation and self-regulation (Pintrich & Schunk, 1995). Bandura hypothesized that self-efficacy beliefs influence other motivational constructs, effort, choice of activities, achievement and persistence. For example, individuals with low self-efficacy for completing a task may choose to avoid it and those who feel efficacious are hypothesized to persist longer and work harder in the face of obstacles (Bandura 1989; Bandura, 1977; Schunk, 1991). Thus, self-efficacy acts as a mediator for performance, academic outcomes and cognitive engagement (Patrick & Hicks, 1997; Hall & Ponton, 2005). Students with high self-efficacy engage is more effective self-regulatory strategies at each level of ability (Bouffard-Bouchard, Parent & Larivèe, 1991). In academic settings, self-efficacy research has been explored extensively in several domains, including science and mathematics where it has been shown to predict outcomes such as academic performance, motivation and other psychosocial constructs (Schunk, 1991). The relationship between self-efficacy and student achievement has been confirmed by several researchers (Hampton & Mason, 2003; Multon et. al, 1991; Pajares & Miller, 1994; Shell et. al, 1995). Studies of college students who pursue science and engineering courses have shown that the academic persistence, necessary to maintain high academic achievement, is influenced by high self-efficacy beliefs (Lent, Brown, & Larkin, 1984, 1986; Pajares, 1997). Academic self-efficacy also correlated with semester and final year grades, in-class homework, exams and quizzes (Pintrich & De Groot, 1990).

Studies conducted in mathematics have shown that college undergraduates' interest in mathematics and their choice of math-related courses and majors is predicted to a greater degree by their mathematics self-efficacy than their prior math achievement (Hackett, 1985; Hackett & Betz, 1989; Lent, Lopez & Bieschke, 1991). Pajares (1996) used a path model to examine the interactions between self-efficacy judgments and mathematical problem-solving of middle school students mainstreamed in algebra classes; in this model, math self-efficacy made a unique contribution to the problem-solving performance of regular education students (r=.387) and of gifted students (r=.455) when the model was controlled for the effects of math anxiety, cognitive ability, mathematics grades, sex and self-efficacy for self-regulatory behaviors. Pajares also reported that girls, including gifted ones, consistently underestimated their confidence even when their scores warranted greater confidence. This gender differential in self-efficacy judgments continued in college, where male undergraduates reported higher mathematics self-efficacy than did female undergraduates (Hackett, 1985; Hackett & Betz, 1989; Lent, Lopez & Bieschke, 1991); this differential is manifested in the negative stereotype that female students have weaker math ability than male students. Thus, the performance of high-achieving female students on challenging math tests can be impaired by a phenomenon known as stereotype threat, which emerges when a negative task-relevant stereotype is activated (Steele & Aronson, 1995). Negative stereotypes about women can lower their performance, self-efficacy, and in combination, these effects can impact women's career decisions. Consequently, the underestimation of confidence rather than lack of skill is cited as one of the primary reasons young female students have exhibited avoidance behaviors towards math-related courses and careers (Hackett, 1995).

Despite the importance of self-efficacy in predicting behavior, it is not solely responsible for behavioral mechanisms. Other variables that come into play, especially in achievement settings, include outcome expectations, skills and the perceived value of outcomes (Schunk, 1991). The lack of skills will result in incompetent performances even if self-efficacy is high; outcome expectations play a key role because individuals usually act in ways they believe will cause positive outcomes; desiring certain outcomes relative to others is the perceived value that people place on outcomes. When these variables (skills, outcome expectations, perceived value of outcomes) manifest themselves in optimal ways, self-efficacy is hypothesized to impact much of human behavior (Bandura, 1989).

Outcome expectations

Outcome expectations are defined as an individual's judgment of the likely consequence of a behavior (Bandura, 1989). They answer the question, "If I do this, what will happen?". People's notions of outcomes can have different sources – symbolic thinking, vicarious experiences and modeling behaviors and the actual incentive value of the outcome (Bandura, 1977; Bandura, 1986). The origin of outcome expectations can be traced back to expectancy-value theories, which emphasize the idea that behavior is jointly impacted by (a) people's perceived expectations of obtaining a particular outcome when performing a behavior and (b) the extent that they value those outcomes (Schunk, 1991). Bandura stated that individuals are more likely to engage in behaviors in which they place greater importance or value on the outcome expectation. However, he ultimately stressed the importance of self-efficacy and noted that the value placed on the outcome expectation is immaterial if the individual does not have the self-efficacy to perform the task and be rewarded (Bandura, 1986; Fouad & Guillen, 2006). Thus, self-efficacy is hypothesized to determine outcome expectations as the expectation of achieving desirable outcomes in a task is tied to an individual's self-efficacy in performing that task (Lent, Brown & Hackett, 1994). While outcome expectations are hypothesized to directly influence interests, intentions and activities (Fouad & Guillen, 2006), Bandura (1997) noted that the dependency of outcome expectations on self-efficacy evaluations will prohibit the former from making a unique contribution to predictions of behavior when self-efficacy perceptions are controlled. Despite studies supporting the construct validity of outcome expectations, it is still an unexplored construct, individually and when modeled with other behavioral constructs.

Hackett and Betz (1981) focused on self-efficacy to explain traditional career choices of females, consequently applying Bandura's SCT to vocational choices. They hypothesized that the limited range of career options for females could be attributed to their low self-efficacy (Betz & Hackett, 1981); this hypothesis was empirically tested using college students and revealed gender differences in their confidence to complete the responsibilities and requirements for nontraditional occupations. Male and female students demonstrated higher self-efficacy levels for traditionally male and female occupations respectively (Betz & Hackett, 1981).

Rather than focus on the type of vocational choices, Hackett (1995) decided to examine the factors that influenced the vocational decision making process. Using the mechanisms of interest development, self-efficacy, outcome expectations and goals and the idea that these person and contextual variables are dynamic interactions, Lent, Brown and Hackett (1994) presented their framework – social cognitive career theory (SCCT) – in a landmark article that set the stage for SCCT to become the most frequently used framework for studying academic and career development.

Social Cognitive Career Theory

The SCCT framework incorporates several environmental and person variables and hypothesizes the manner in which these variables interact to affect an individual's career interests and behavior. Three explanatory models, each constituting different sociocognitive mechanisms, form the core of the SCCT framework. These models were developed to understand the mechanisms by which (a) career and academic interests are formed – the interest development model, (b) career choices are realized – model of career choice, and (c) career performance outcomes are achieved – model of performance (Lent, Brown & Hackett, 1994, p.80).

SCCT model of interest development

As shown in the center of **Figure 2.1**, this model links self-efficacy and outcome expectations to the development of occupational interests. Over the course of childhood and adolescence, people are directly and vicariously exposed to several occupationally relevant activities in their environments (Lent et al., 1994). Differential reinforcements received for continued engagement in different activity domains influence people's self-efficacy beliefs (Bandura, 1986).

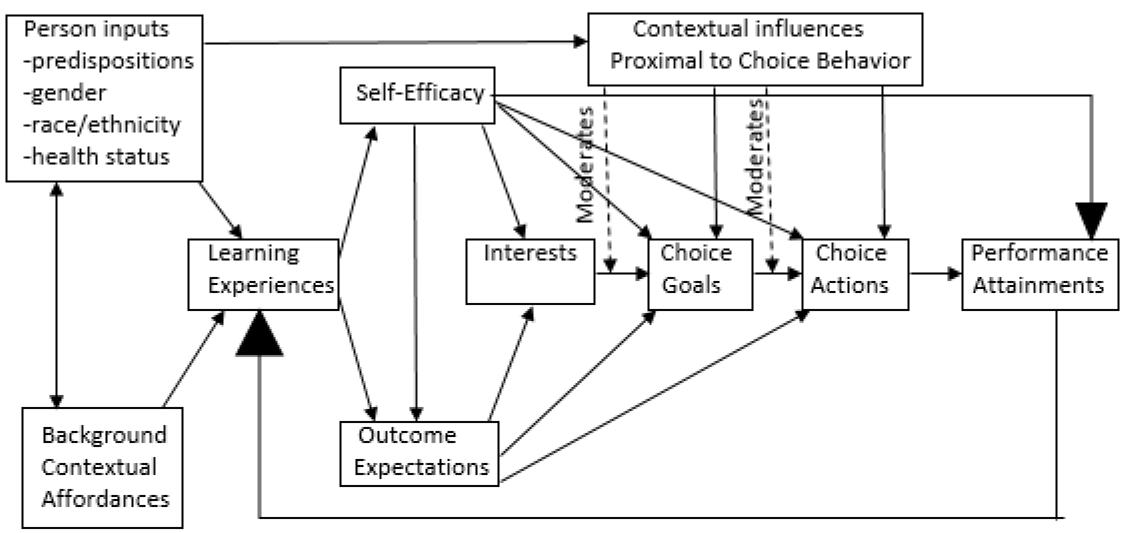


Figure 2.1. Social Cognitive Career Theory Model (Lent, Brown & Hackett, 1994, Figure 2, p. 93).

Lent et al. (1994) hypothesized that these people are most likely to develop interests in activities in which they are efficacious and from which they expect positive outcomes. An ongoing feedback loop is created in which sustained involvement in an activity leads to subsequent mastery or failure experiences, which help revise self-efficacy beliefs, outcome expectations and ultimately interests. Development of interests demonstrates fluidity until late adolescence, the point at which these interests stabilize. However, exposure to new learning experiences such as parenting or job training in post adolescent years can alter an individual's sense of self-efficacy and outcome expectations, thereby changing their interests (Lent et al., 1994). Regardless of objective talent, the formation of strong self-efficacy beliefs and positive outcome expectations are essential in the development of interests.

SCCT model of career choice

While this model builds on the model of interest, the key distinction is that in the model of career choice, career-related interests are linked to goals and actions, especially related to occupational decisions. The choice model emphasizes that learning experiences give rise to self-efficacy beliefs and outcome expectations and these experiences are influenced by environmental factors such as levels of support, barriers and opportunities afforded to a person (Lent et al., 1994). SCCT hypothesizes that when contextual factors moderate the formation of choice goals and execution of choice actions, interests will be a strong predictor of the types of choices people make depending on the environmental conditions. Under supportive conditions, interests are expected to have the greatest influence on academic and occupational choices. On the contrary, restrictive conditions may require individuals to compromise their interests and consider the more culturally acceptable or pragmatic choice (Lent et al., 1994). As depicted in **Figure 2.1**, a feedback loop is developed between performance attainments and learning experiences.

SCCT model of task performance

The performance model, a subset of the career choice model, links self-efficacy and outcome expectations to performance goals, which then lead to performance attainment levels (Lent et al., 1994). This model is concerned with predicting and explaining two primary aspects of performance: their accomplishments and behavioral persistence (e.g. stability in an academic major). In this model, as shown in **Figure 2.2**, abilities and past performance inform self-efficacy beliefs and outcome expectations (Lent et al., 1994). Self-efficacy and outcome expectations work in conjunction with ability, in part by influencing the types of performance goals. When ability levels are controlled, high self-efficacy and positive outcome expectations dictate the performance goals individuals establish for themselves. This model excludes interests as a mediating variable; according to Lent et al., (1994) interest are "more integral to choice of career/academic activities than to selection of performance goals" (p. 99).

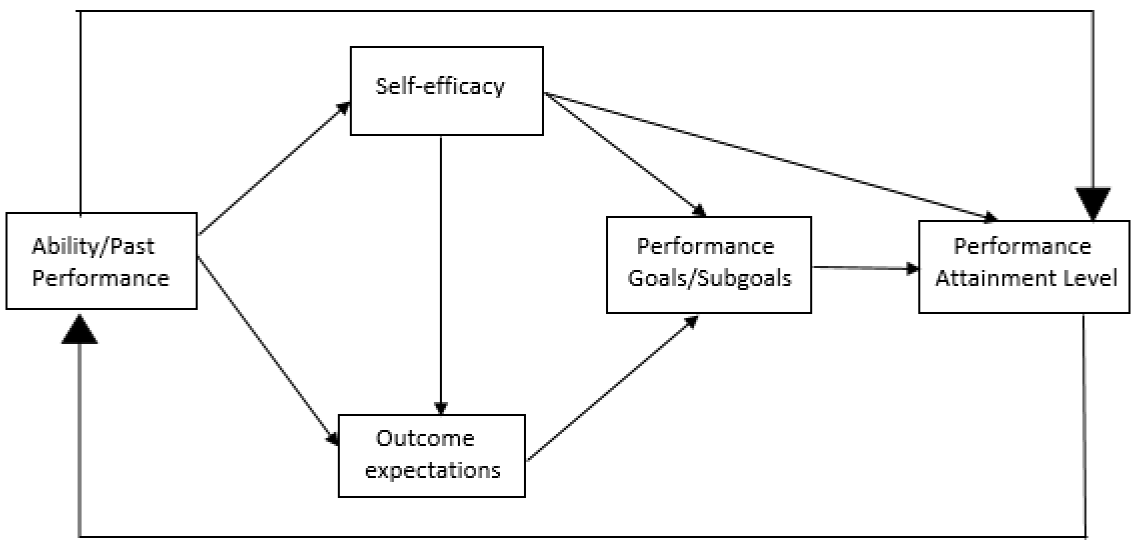


Figure 2.2. SCCT Model of task performance (Lent, Brown & Hackett, 1994, Figure 3, p. 99).

SCCT in STEM

As SCCT has been a useful framework for explaining educational and vocational interests, choices and performance, a substantial body of research has been conducted to test the theorized relationships in each SCCT model.

The model of interest development has been explored in math, in which Lent, Brown and Hackett (1994) posited that math self-efficacy was significantly and strongly correlated with math interests (*r* range=.53-.63). Using path analysis, Smith and Fouad (1996) tested several hypotheses from Lent et al.'s work (1994) in a study using middle school students who had similar socioeconomic status (SES). Results showed strong positive correlations between self-efficacy and outcome expectations (*r*=.55) and between outcome expectations and intentions (*r*=.39). Self-efficacy had a strong direct influence on interests (*r*=.29), which subsequently influenced intentions (*r*=.28).

When examining predictors of goals, Waller (2002) and Lent et al., (1993) conducted studies that showed significant correlations between math self-efficacy and math goals (*r*=.46 and *r*=.63 respectively), math outcome expectations and goals (*r*=.42 and *r*=.52 respectively) and math interests and goals (*r*=.68 and *r*=.71 respectively). Lent et al., (2008) tested the predictive utility of the SCCT model using a diverse sample of students majoring in computing disciplines. Results showed that the SCCT model accounted for nearly 40% of the variance in interests and 33% of the variance in persistence goals for these students.

While hypotheses in the interest and choice models have been tested using all SCCT predictors, the same cannot be said about the performance model as most studies have usually examined only subsets of this model. Meta-analyses results have indicated correlations of .38 - .50 between self-efficacy and college student academic performance (Multon et al., 1991; Robbins et al., 2004). A full corrected correlation of .36 was also reported between academic self-efficacy beliefs and college retention criteria. Finally, as hypothesized by SCCT, a significant correlation of *r*=.45 was observed between indices of past (high school) academic performance and college performance (Robbins et al., 2004).

Progressing beyond bivariate correlations, Brown et al., (2008) used path analyses to model academic performance and academic persistence. General cognitive ability (e.g. ACT or SAT scores) and past performance (high school GPA) were used as predictors respectively. As meta-analytic estimates were unavailable for outcome expectations, the researchers conducted their own meta-analyses of outcome expectations to fill in the gaps. However, as only self-efficacy-outcome expectations relationships had been sufficiently examined to provide valid meta-analytic estimates, outcome expectations were excluded from the analyses and a reduced version of SCCT's performance model was tested (Brown, 2008). Results indicated that self-efficacy seemed to influence academic performance directly as opposed to being mediated by goal mechanisms. This finding was not aligned with SCCT's model of academic performance perhaps due to way goals were operationalized; given that goals were measured as intentions to complete college rather than as performance indicators, there was a mismatch between the measurement and the outcome it was predicting (Brown et al., 2008). However, goals were much better predictors of retention suggesting that the reliability of social cognitive variables is dependent on how well they match criterion variables (Lent & Brown, 2006).

Brown et al., (2008) also found that indices of academic aptitude showed indirect relationships to college retention outcomes via self-efficacy beliefs and goals to complete college. Thus, for students with similar academic abilities, the likelihood of finishing college was influenced by the confidence they placed in their academic abilities and the goals developed for college completion (Brown et al., 2008). Between the two predictors used for this study, prior high school performance showed a strong relation to self-efficacy beliefs in comparison to general cognitive ability; however, cognitive ability rather than past high school performance seemed to inform college performance to a greater degree (Brown et al., 2008). Despite these findings, the generalizability of SCCT has been limited due to the predictors and outcome variables in the models being examined at a single time point, resulting in cross-sectional studies. Additionally, these models utilize regression and correlational analyses to assess relationships among variables; while these techniques are beneficial in supporting SCCT's hypothesized relationships among its constructs, the cross-sectional nature of the models does not allow for causal pathways to be established between the predictors and criterion variables (Lent, 2008).

Using a sample of mostly first semester engineering students, Lent et al., (2008) extended the longitudinal study of SCCT's interest and choice models by examining variables besides self-efficacy and interests; in addition, this was conducted at two points in time, five months apart to explore the temporal relations among the variables. Three possible models were proposed for the role of self-efficacy: antecedent, consequent and bidirectional. The antecedent model conceived self-efficacy as a precursor to outcome expectations, interests and goals, the consequent model viewed it as a consequent of the other variables and the bidirectional model conceptualized self-efficacy as having a reciprocal relationship with the other variables (Lent et al., 2008). Results from this study indicated that the antecedent model provides "a sufficient and parsimonious explanation of the relations among the theoretical variables". The predominant temporal flow appeared to be from self-efficacy to the other variables rather than vice versa" (Lent et al., 2008, p.333).

SCCT in Studies with Underrepresented Populations

Given that SCT was extended into SCCT based on the idea that low self-efficacy deters females from selecting careers of their choice, several studies have examined the role of sex in the context of each SCCT model. In addition, the influence of marginalized groups within each SCCT model have also been investigated.

Using a scale created to measure educational and career-related barriers in high school students, McWhirter (1997) found that male and European American high school students experienced fewer educational and vocational barriers in comparison to female and Mexican-American high school students. However, in a study that used SCCT to predict interests and major choice for male and female students in engineering, Lent et al., (2005) found that although there were no differences between male and female students in terms of their self-efficacy, outcome expectations or interests, female students encountered fewer social barriers and experienced more social support in their pursuit of an engineering major than did male students.

To investigate the contextual factors from SCCT comprehensively, Fouad et al., (2010) developed an instrument to recognize the vocational and educational barriers and supports perceived by male and female students in the mathematics and sciences domains; this study was conducted at three educational levels: middle school, high school and college. The results from this study are summarized in **Tables 2.1** and **2.2** respectively. **Table 2.1** displays the results for the top supports and barriers in math while **Table 2.2** displays the top supports and barriers in science.

Table 2.1. Top supports and barriers in Math for male and female students at three educational levels (Fouad et al., 2010)

Top supports in Math	Middle School	High School	College
Males	Teachers	Clarity in career goals	Teachers
Females	Teachers	Teachers	Teachers

Top barriers in Math	Middle School	High School	College
Males	Lack of role models	Uninspired teachers	No opportunities outside school
Females	Teachers	Teachers	Teachers

Table 2.2. Top supports and barriers in Science for male and female students at three educational levels (Fouad et al., 2010)

Top supports in Science	Middle School	High School	College
Males	Teachers with high expectations	Teachers with high expectations	Inherent interest in subject
Females	Teachers with high expectations	Teachers with high expectations	Inherent interest in subject

Top barriers in Science	Middle School	High School	College
Males	Uninterested friends	Uninspired teachers	Lack of help from parents
Females	Uninspired teachers	Test anxiety	Uninspired teachers

In a study conducted in Spain, Inda et al., (2013) used sophomore engineering students to test SCCT and the role of gender in predicting engineering interests and major choice goals. Findings indicated that female students have weaker self-efficacy beliefs and interest than male students, despite the lack of differences in outcome expectations and goals. In addition, peers and parents are among the top support systems for female students while male students perceive more parental barriers than females (Inda et al., 2013).

From a broader perspective, the findings of these studies strengthen the utility of the SCCT model in the context of career development. The affective domain addressed in science and engineering studies is no different than what instructors and advisors encounter in chemistry, especially in gateway courses, which serve as crucial points when students decide to persist in or change their academic paths. The low self-efficacy beliefs and interest demonstrated by female students in engineering have already been documented in chemistry; Zusho et al., (2003) observed decreased motivational levels among students, especially low achievers, across a semester in introductory chemistry. In organic chemistry, males have reported higher levels of self-efficacy than female students (Lynch & Trujillo, 2010). In recent years, several statistical models have been tested and used to predict student achievement in college chemistry. In an effort to examine meaningful learning in chemistry, Brandriet et al., (2013) used structural equation modeling (SEM) to test the relationship between cognition, affect and chemistry achievement. Results showed the existence of a tripartite relationship among the three variables. By using students' math ability, prior conceptual knowledge in chemistry and attitude towards chemistry as predictors, Xu, Villfane and Lewis (2013) used SEM to predict achievement in chemistry; results showed that the three predictors accounted for 69% of the variance in chemistry achievement. Given the domain specific nature of affective constructs and the nuanced yet sometimes conflicting gender differences in various domains, including chemistry, it is essential to examine the impact of affective and cognitive variables on students' performance and persistence in their intended STEM majors within the context of chemistry gateway courses.

While much of the literature has utilized SCCT in cross-sectional and longitudinal models to support hypothesized relationships and paths in the framework, these models have been incomplete due to the lack of operationalized outcome expectation measures. This exploratory study aims to fill this gap by developing an instrument to measure outcome expectations in chemistry. This instrument, in conjunction with self-efficacy and measures of cognitive ability, will be used to develop a comprehensive model of performance and STEM persistence in gateway courses populated by STEM majors as opposed to chemistry majors. This complete model and examination of its affective components on a finer scale are expected to offer a much better understanding of how students persist in a STEM major, when they might leave this major and how to intervene and potentially remediate this situation.

CHAPTER 3: INSTRUMENTS AND METHODS

The development of a model that combines persistence and performance requires the use of psychometrically sound persistence measures to capture latent constructs (unobserved variables). This chapter provides a brief description of the full-length, norm-referenced, self-report quantitative research instruments that were adapted or developed to measure self-efficacy and outcome expectations. Some information about a subset instrument – developed from the full-length surveys and used to collect data at several points during a semester – is provided as well. Also included are the research design, a general framework for scale development, and descriptions of samples to which surveys were administered. Only the methods common to development and psychometric evaluation of the two full-length surveys will be described here.

Norm-referenced instruments

Norm-referenced instruments serve multiple purposes; they can be utilized to classify students, assess progressive changes or predict results of some tests. These instruments reveal differences between and among students based on the characteristics being measured and establish a rank order of students across a continuum of values (Mishel, 1998; Waltz et al., 1991; Pett et al., 2003; Bond, 1996).

In this case, the instruments used to measure self-efficacy and outcome expectations were designed using norm-referenced frameworks as the surveys were measuring specific characteristics and the goal of each survey was to obtain a range of students' scores that would enable the researcher to discriminate one student's self-efficacy or expectations from those of other students or a norm group (Pett et al., 2003).

Instruments and participants

The instruments, their target populations and modes of administration are summarized in **Table 3.1**. Instrument development, scale construction and psychometrics will be discussed further in the chapters dedicated to each instrument.

Table 3.1: List of instruments and administration details

	Instrument	Construct(s)	No. of items	Mode	Timing of administration	Target populations
1	Chemistry self-efficacy and anxiety survey (CSEAS)	Self-efficacy and anxiety	30	Pilot (paper), electronic	Start and end of semester (pre-post)	Preparatory chemistry, general chemistry I and II and chemistry for engineering majors
2	Chemistry outcome expectations survey (COES)	Outcome expectations	25	Paper	Start and end of semester (pre-post)	Preparatory chemistry, general chemistry I and II and chemistry for engineering majors
3	Subset survey (combination of CSEAS and COES)	Self-efficacy and outcome expectations	25 (13 from CSEAS & 12 from COES)	Paper	Before and after each hourly exam during the semester (exams 1,2 & 3)	General chemistry I and II

Mixed methods design

The studies detailed in chapters 4 – 6 involve the collection of data using quantitative survey instruments. Given the latent nature of the constructs being measured, quantitative research is often insufficient to understand the context or setting in which students respond (Creswell, 2003). In addition, the inability to actually communicate with students and hear their opinions and biases can cloud the interpretation of these surveys. Although survey construction usually starts with qualitative data collection, in this study quantitative data collection was followed by qualitative research in the form of semi-structured interviews. While these interviews offer rich data that aid in scale development, the limited number of participants restricts generalizability of these findings. Thus, quantitative and qualitative data collected in combination or in sequence offer more thorough evidence for understanding the research questions posed in this dissertation. This justifies the use of sequential exploratory mixed methods research in collection and analyses of quantitative and qualitative data (Creswell, 2003).

Pilot versions of the CSEAS and COES – containing open- and close-ended items – were administered to students for the purposes of scale development and coding of open-ended responses for subsequent analyses. Following the initial administration, semi-structured student interviews were conducted to elicit interpretations of each survey and consequently refine survey items. A second round of these interviews was conducted after finalized versions of each survey were in administration. Comparing the data to evaluate similar patterns obtained subjectively and statistically allowed for data triangulation and psychometric support that would have been difficult to establish using either a qualitative or quantitative approach alone (Towns, 2008).

Description of samples

Preparatory chemistry

This 4-credit course serves as an introductory course in general inorganic chemistry intended for students with little or no previous science background. Acting as a feeder course for traditional and returning students, this course constitutes students with a variety of majors; following completion of preparatory chemistry, students typically enroll in GCI, II, or general chemistry for engineers. This course has discussions but does not have an associated laboratory component.

General Chemistry I and II (GC I and II)

These courses (5 credits each) form the two-semester gateway sequence of introductory college chemistry courses. Enrolled students include all science majors (except nursing and engineering) and some allied health majors usually planning to enter professional programs in medicine, dental hygiene, pharmacy, physical therapy and the like. These courses also administer an internal placement test to assess students' backgrounds in mathematics and chemistry. General Chemistry I (GC I) administers the Toledo Chemistry Placement Exam (TCPE), written by the American Chemical Society (ACS) while General Chemistry II (GC II) offers one component (paired exam – GC05PQF) of the two-part standardized final exam taken by students in GC I. Both courses have associated discussion and laboratory components. Both courses also take an ACS standardized final exam (GC05PQF and GC08C).

General Chemistry for Engineers

This is the terminal course in chemistry for engineering majors. Composed primarily of engineering majors, this course is also taken by some finance, math and computer science majors. This course has an associated discussion and laboratory component. Students take a standardized final exam.

General scale development

This section describes a general framework for instrument development. While this framework served as a guide to the development of both surveys in this dissertation, each survey had specific criteria that needed to be fulfilled; these issues will be discussed in detail in the chapters dedicated to each survey.

a) The first step in instrument development requires defining the target construct accurately followed by assessing the need for instruments to measure these constructs.

b) The type of scale and response style for the instrument was determined based on the analyses to be conducted on the data. The need to measure latent variables, obtain descriptive statistics and subsequently analyze resulting data justified the use of a subject-centered Likert-type scale in the design of both instruments (Dawis, 1987) in this study. The Likert-type scales described here are considered 'quasi-interval' – a type of scaling that falls between ordinal and interval. While technically ordinal, these scales are the sum of Likert-type items comprising of five or more levels of the latent variable being examined and as such end up being a reasonable approximation of an interval data point (Rattray & Jones, 2007).

c) An initial pool of items was generated from an extensive literature search for existing measures after which the number of items was reduced based on consultations with principal investigator and fellow researchers.

d) The type of items, order in which they were presented and the language used were important decisions to make in order to avoid biased responses. Open ended questions were added, if necessary.

e) Surveys were pilot tested with each sample of intended respondents.

f) Psychometric analyses were conducted on each survey. The methods constituting psychometric evaluation are discussed later in this chapter.

g) Following pilot testing and item deletion, a final version of survey was administered to a different sample of respondents with ongoing psychometric testing.

Brief description of instruments

The first instrument, focused on measuring students' perceived strength of self-efficacy beliefs, is the Chemistry self-efficacy and anxiety survey (CSEAS) – in administration since Fall 2012 and comprising of 30 items measuring self-efficacy and 15 items measuring anxiety. Items

/ statements in this survey were vetted with students in semi-structured interviews. Each item was evaluated on a Likert scale of 1-6 (1 = not confident at all to 5 = totally confident and 6 = not applicable / not sure), adapted from Coll (Coll, Dalgety & Salter, 2002), and the survey was administered in a pre/post manner. A short anxiety survey was also integrated within the CSEAS to collect data that would aid in establishing some degree of validity. In addition, the instrument asks for students' majors and includes a plethora of questions to assess student behaviors and interests. As this survey is offered during the first week of class, the majors indicated in the CSEAS are more current than the information provided through institutional research; consequently, these majors are coded and used for any analyses that requires this information. Demographic information for the students, details about the instrument development process and establishment of psychometric reliability and validity from the resulting data are presented in chapter 4.

The second instrument, focused on measuring students' outcome expectations, is the Chemistry outcome expectations survey (COES) – in administration since Fall 2013. Given the scarcity of surveys that capture outcome expectations, especially in chemistry, this instrument would be the first in chemical education research to measure this construct. Students were asked to indicate their level of agreement, with several 'if-then' statements (Fouad & Guillen, 2006) using a Likert scale of 1-5 (1 = strongly agree and 5 = strongly disagree). The 'if' sub-statement was associated with a particular task; 'then' was used to phrase the outcome sub-statement (e.g. If I work hard enough, I will pass this course). Comprising of 25 items, this survey was also administered in a pre/post manner. Demographic information for the students, details about the instrument development process and establishment of psychometric reliability and validity from the resulting data are presented in chapter 5.

The third survey was a subset instrument – in administration since Fall 2014 – comprising of items from both the CSEAS and COES. While the utility of the full length CSEAS and COES was tested on students in preparatory chemistry, GC I, GC II and chemistry for engineers, the subset instrument was only administered to students in GC I and II as these courses constitute the two-semester gateway sequence and serve as important decision points in the persistence model. This instrument consisted of 25 items from both persistence measures (13 self-efficacy items from the CSEAS and 12 items from the COES) and used the same Likert scales as the original full length instrument for each construct. This instrument was also vetted with students in semi-structured interviews. Demographic information for the students, details about the instrument development process and establishment of psychometric reliability and validity from the resulting data are presented in chapter 6.

Instruments that were administered on paper ("fill in the bubble" forms) were scanned and processed using Remark Classic OMR (Optical Mark Recognition) software; the resulting data were saved as Excel files for subsequent screening and use in analyses. Instruments administered online were done so using Qualtrics; parameters of the survey included forced response for all items, time stamps for total time taken, headers repeated at page breaks, page separation for different components of the survey and the inability to go backwards. Resulting data were exported into Excel accordingly. In addition, for the entire length of the study, demographic data was sought through institutional research to offset any possible stereotype threat (Steele & Aronson, 1995; Steele, 1997). Stereotype threat is a concern that members of underrepresented groups experience about their performance or actions reinforcing or confirming a negative group stereotype (Steele, 1997). For instance, the pervasive negative stereotype about "boys being better than girls in mathematics" can result in girls performing poorly on a test they believe is measuring math ability due to their anxieties about confirming the negative stereotype (Spencer, Steele, & Quinn, 1999). In many cases, the process of identifying oneself as female before taking a math test was sufficient to trigger anxiety and result in lower test scores (Danaher & Crandall, 2008). Consequently, instead of having students provide demographic information as part of survey completion, these data, which included students' sex, ACT scores, intended major and minor, high school GPA and educational level, were obtained through institutional research. Some of this information was utilized for characterizing the sample and for analyses relevant to each study.

Methods and data analyses

The following statistical procedures constituted the analyses that were completed to characterize the sample, evaluate the instruments and assess the relationships among the variables that framed each study.

Descriptive statistics

Given the considerable number of variables in each study, descriptive statistics were useful in summarizing these variables and presenting this information in a manageable form. Measures of central tendency like mean, median and mode provide the most basic information about the observations in a data set. In addition, assessing the dispersion and normality of distributions also offers a better understanding of the sample being studied. Descriptive statistics examined in this work include cross-tabulations, means, standard deviations, skewness and kurtosis values. Normality of distributions are assessed as required by the statistical technique.

Comparative statistics

T-tests and associated effect sizes were used to make comparisons between groups of students. Paired sample t-tests were used to examine differences between measurements from the same group of students. Given the pre-post administration of surveys and comparisons in a single group, these tests were useful in detecting statistical differences between the means in a population.

Independent sample t-tests were used to examine differences between two unrelated groups. These tests were useful in evaluating changes between subgroups based on gender, ability (high vs. low), performance (high vs. low) or construct (high vs. low self-efficacy or outcome expectations).

Effect sizes, measured by Cohen's d, were reported in both tests to evaluate the magnitude of mean differences between groups (Cohen, 1988).

Correlation analysis

A bivariate correlation describes the extent of a relationship between two variables. The correlation coefficient '*r*' is a single value that expresses the direction and degree of linear relationships between two individual variables in a sample. Depending on the type of data being analyzed, the correlation coefficient can be a Pearson product-moment correlation, Spearman Rank, Lagged and others. Correlation analyses were used in aspects of psychometric testing for the CSEAS and COES. When multiple variables are involved, the correlations between each pair of variables is expressed in the form of a correlation matrix (Thorndike, 1978).

This matrix serves as the starting point for various statistical techniques, including those used to analyze data resulting from the studies described in this dissertation. Correlation analysis was utilized in two basic ways: to reveal relationships between variables for informational and decision making purposes and to determine the predictive ability of a variable such as in the case of regression analysis, which will be discussed in some detail in chapter 6. As with any other statistic, the proper interpretation of correlation coefficients depends on the sampling scheme used to generate the data. Adequate heterogeneity in sampled groups is essential to allow for the manifestations of relationships (Miles & Banyard, 2007).

Factor analysis – Exploratory

Factor analysis refers to a collection of statistical techniques used to examine relationships among complex variables. When these variables are latent – such as psychological or affective states - and there is no direct method to measure them, surveys incorporating multiple items are developed with the idea that there are underlying unobservable factors that will emerge based on patterns in the survey responses (Field, 2009). The goal of factor analysis is to assess the patterns of responses and regroup a large set of variables into smaller sets of factors; each factor is an indication of the overall variance in the observed responses (Yong & Pearce, 2013). In this study, factor analysis was used in psychometric instrument development as a way to refine the item pool for each Likert-scale survey; in addition, it also contributed towards establishing aspects of psychometric testing. The first phase of factor analysis was purely exploratory, aptly name exploratory factor analysis (EFA). Once an underlying structure was established for each survey, a confirmatory factor analysis (CFA) was used to test the stability of this structure.

There were several decisions involved in conducting factor analyses and interpreting the results. The following steps describe these decisions, starting with some checks that were necessary even before analysis was conducted:

a) <u>Adequacy of sample size</u> – The guidelines for adequacy of total sample size are varied. While Tabachnick and Fidell (2001) recommend a sample size of at least 300 cases, they also indicate that smaller sample sizes (e.g., $n \approx 150$) are adequate when other criteria necessary for accurate interpretation of factor analysis results are satisfied. A minimum sample size of 100 cases is recommended provided other requirements meet necessary standards (Rattray & Jones, 2007).

These guidelines have been criticized by some researchers who suggest that the data collected should dictate the appropriate sample size; thus, the aim should be to obtain the largest possible sample and make judgements about the adequacy of this sample size post data analysis (Henson & Roberts, 2006). In this study, homogeneous data sets such as GC I-post and GC II-pre from the same semester were combined once t-tests indicated that students in both groups had similar ability levels (no significant differences) based on their ACT or TP scores.

b) <u>Fulfilling assumptions</u> – Data collected should be from a random sample and fulfil the assumptions required of multivariate statistical techniques namely absence of outliers, linearity, continuous data and low percentage of missing data (Comrey, 1985; Pett et al., 2003; Beavers et al., 2013; Child, 2006). Multivariate normality is an assumption depending on the method used to extract factors (Stevens, 2002).

c) <u>Assessing the correlation matrix</u> – The resulting correlation matrix was examined for moderate to strong correlations among the variables (survey responses) because these are essential for patterns to emerge and result in factors. The correlation matrix was also evaluated for singularity based on its determinant. The determinant indicates whether the vectors comprising the matrix are linearly independent. If rows or columns in a matrix are zeros, equal or a linear combination of other rows or columns, resulting in linear dependencies, the determinant of this matrix will be zero and this indicates singularity. In factor analysis, a determinant close to zero would indicate presence of potentially redundant items that are strongly correlated and deserve further evaluation (Pett et al., 2003). Other indices provided by the statistical software were also used to assess the appropriateness of the correlation matrix: Kaiser-Meyer-Olkin (KMO) which measures the shared variance in the items, Bartlett's test of sphericity which tests the null hypothesis that the correlation matrix is singular and measures of sample adequacy (MSA) which examine the correlations of an item with other items in the matrix (Pett et al., 2003).

d) Extraction of factors – A variety of methods are available for fitting the factor analysis model. Principal Component Analysis (PCA) is used in data reduction and reduces a large number of items to smaller components. The goal of PCA is parsimony wherein the maximum amount of variance is extracted using the smallest number of factors (Field, 2009). In addition, PCA extracts components that include the total variance (common, unique and error). Other extraction methods include principal axis factoring (PAF) and maximum likelihood. PAF parses out unique and error variances, thus accounting for just common variance among factors while maximum likelihood not only extracts factors but also provides additional information such as fit statistics (Pett et al., 2003). In this study, both PCA and PAF were used to extract factors, make comparisons between the resulting factor structures and decide which structure was meaningful substantively and statistically. If these techniques produced highly dissimilar structures, this would call for reevaluation of the data – the correlation matrix would be examined for low off-diagonal elements which could result in low communalities (proportion of each variable's variance that can be explained by the component) (Field, 2009).

e) Retention of factors – Several criteria were followed when deciding on how many factors to retain. Mathematically, there can be as many factors as there are variables; however, not all these factors contribute to the overall structure (Henson & Roberts, 2006). As the goal was to explain the largest variance using fewest factors, only meaningful factors explaining aspects of the construct being examined were retained. Kaiser's criterion retains factors that have eigenvalues greater than 1. Given the overestimation of factors using this criterion, the scree plot was also used to make determinations about factor retention. The number of factors above the natural bend or 'elbow' were retained (Costello & Osborne, 2005; Thompson & Daniel, 1996). Parallel analysis – a Monte Carlo simulation technique – offered the best approach to determine how many factors to retain. This method makes comparisons between the magnitude of eigenvalues obtained using the dataset in question and those obtained from randomly generated correlation matrices of the same size; factors retained were dictated by the number of eigenvalues (generated from the researcher's dataset) that were larger than the corresponding random eigenvalues (Horn 1965). The parallel analysis engine used to calculate these eigenvalues was available online (Patil et al., 2007; Patil et al., 2008; O'Connor, 2000). In this dissertation, a combination of criteria was used to decide how many factors to retain. In addition to those mentioned above, the residuals, percentage of cumulative variance and characteristics of the resulting factor structure were evaluated to make this decision.

f) Factor rotation – Rotating each group of items toward the axis allows for easy interpretation of the factor structure (Osborne, 2015). Two commonly used rotation methods are orthogonal – where the factors are uncorrelated to each other – and oblique, in which the factors are correlated. Achieving 'simple structure' was the goal behind selecting a certain type of rotation. Simple structure is a condition in which each item has a high or important loading (absolute value near 1) on one factor only, each factor has meaningful loadings for only some of the items and no variable cross loads (Pett et al., 2013). While it is expected that most factors would correlate with each other, thus justifying use of an oblique rotation, the main concern of these exploratory analyses was to identify meaningful dimensions resulting from structuring of variables. In addition, the methods used in much of the literature dedicated to instruments developed here utilized an orthogonal rotation to achieve simple structure. Due to these reasons, Varimax rotation, the default orthogonal rotation in SPSS, was used in the factor analyses conducted on each instrument developed in this dissertation. As long as simple structure is clear, either method of rotation should result in similar interpretations (Kline, 2002).

g) Interpretation – Each factor structure was examined for a few criteria: Mathematically and substantively, a factor was considered meaningful only if it had three or more constituent survey items, each with a strong correlation ("loading") with the factor. As the analyses in chapters 4 and 5 were mainly based on principal components (but checked with principal axis factoring), resulting in higher estimated loadings, absolute loadings less than 0.50 were suppressed; this was a stringent criterion indicating that 25% (or more) of the variance on the item was shared with the factor. Items that cross-loaded (appeared in more than one factor) were considered for removal if the loading in each factor was greater than 0.40 (Costello & Osborne, 2005). The process of obtaining the most stable and meaningful factor structure was an iterative one; deletion or modification of items was always ensued by an EFA with a range of factors extracted to allow for comparisons between factor structures. If the items within each factor showed a high degree of relatedness, the scores (survey responses) of these items were combined into a single average subscale score. This process was followed for the factors in both the CSEAS and COES. Obtaining a single, pure factor structure (with sensibly grouped items) for each survey was the first step towards making meaningful measurements longitudinally.

Factor analysis – Confirmatory

Having developed an instrument and obtained a factor structure using exploratory methods, the next step was to confirm the stability of this structure across other population samples. Confirmatory factor analysis (CFA) was used to test the factor structure(s) obtained using EFA.

CFA utilizes a variance-covariance matrix to test a measurement model; as CFA was being used to test the factor structure identified through EFA, the datasets used for CFA were different than those used for EFA (Pett et al., 2003). In addition, as the factor structures from both the CSEAS and COES had to be stable for longitudinal measurements, each structure had to demonstrate a reasonable to good model fit at three different time points – start of GC I, end of GC I and start of GC II. Consequently, data sets from each of these time points were utilized in CFA testing.

The *proc calis* procedure in SAS 9.3 was used to conduct a CFA; code for this procedure is shown in **Appendix A**. While this code shows items from the COES, a similar code was written for CFA conducted using CSEAS items. IBM SPSS AMOS 24 was used to create the CFA path diagram and verify the results from SAS 9.3. For both instruments, the factor model was specified using the latent factors and their constituent variables from EFA. The nonzero loadings (on variables comprising each factor) were designated as free parameters and each observed variable loaded on exactly one factor. Factor variances were fixed at 1.0 and error variances of the observed variables were left as free parameters. The model was tested using correlated factors (co-variances among factors were free parameters) and Lagrange Multiplier (LM) tests offered suggestions on how to improve fit indices for the model. Residuals, outliers and leverage values were also requested for in the SAS code. Models were tested with and without cases that were outliers or had high leverage and residual values to check for improvements. Maximum likelihood (ML) estimation was used as the model-fitting procedure and observations with missing values were excluded. Although CFA requires a much larger sample size than an EFA and the full information maximum likelihood (FIML) method allows inclusion of observations with random missing values, this method was not used here mainly to stay consistent with the deletion procedures that were used in all other analyses.

Model fit was assessed using several types of fit indices: Absolute fit indices, relative fit indices and non-centrality fit measures (Tanaka, 1993; Maruyama, 1998). Absolute fit indices:

a) The chi square test (χ^2) – A non-significant ($p > .05$) chi-square value indicates that the model is acceptable; the observed and predicted covariance matrices are similar. However, chi square tests are highly dependent on sample size, with large samples sometimes resulting in model rejection (Schmitt, 2011). Thus, alternate fit indices are considered when reporting model fit.

b) Standardized root mean square residual (SRMR) – This represents the square root of the mean of the covariance residuals. A good fit is indicated by a value less than .08 (Hu & Bentler, 1999).

Relative fit indices: These indices compare the target model to a null or baseline model.

a) Tucker-Lewis index (TLI) – Values above .95 indicate good fitting models (Hu & Bentler, 1999).

Non-centrality fit indices:

a) Root mean square error of approximation (RMSEA) – Currently the most popular measure of model fit, the thresholds for acceptance are varied. However, according to Hu & Bentler (1999), a value less than or equal to .06 indicates good model fit. In addition, the lower value of the 90% confidence interval should be close to zero and the upper value should be less than .08. Reporting the value of this index along with its confidence interval provides precise information about the estimate of the RMSEA (Hu & Bentler, 1999).

In the instruments developed here, fit indices, parameter estimates, standard errors, standardized residuals and factor correlations, squared multiple correlations were examined to justify fair to good model fit of the factor structures. If these measures indicated poor model fit, modification indices were examined to make alterations to the factor structure; if this did not improve model fit, EFA was conducted again.

Cluster analysis

Cluster analysis is a technique similar to factor analysis and ordinarily used to group people instead of variables. These clusters result from similarity in people's responses to variables or items in a survey. However, for the instruments developed here, cluster analysis was used as an alternative to factor analysis; thus, the goal was to cluster variables (survey items) that were similar to one another. This method was used to obtain factor or 'cluster' structures in two chemistry courses: preparatory chemistry and general chemistry for engineers. SPSS version 22/23 and Excel version 2014/2015 were used to perform these analyses. The rationale for using this technique for these courses was threefold:

a) To test the utility of a method typically reserved for grouping cases and examine the resulting cluster structures.
b) Given that preparatory chemistry and general chemistry for engineers are 'feeder' and 'terminal' courses respectively, the fairly heterogeneous make-up of these courses and resulting heterogeneity in survey responses called for a technique that would allow leeway for non-normality in data and utilize measures of similarity other than the correlation coefficient to analyze the variables.
c) To find an empirical classification and contextual similarity in responses based on the a priori theoretically defined factor structures from general chemistry.

Cluster analysis used the Euclidean distance, *d*, as a measure of similarity with smaller distances between variables representing greater similarity in variables. As this measure is highly sensitive to large variances in responses, variables were standardized before analysis. The 'hierarchical cluster' method – in which variables start out in one cluster and gradually form individual clusters – was used in this study (Everitt et al., 2001). Average (between-group) linkages, based on average Euclidean distance, are used to create the clusters. The resulting dendrogram displays the links between variables and allows for identification of variables that form distinct clusters. As this method was an alternative to factor analysis, the dendrogram was used to create the analogous "rotated factor structure". Cluster and factor analyses were used as complements to examine similarity in association and context when the surveys were administered to students in courses related to general chemistry (Gorman & Primavera, 1983).

Psychometric theory

One of the crucial aspects of survey development involves the psychometric evaluation of data resulting from these surveys. The ongoing validity and reliability testing of these data dictate the utility of these survey instruments in research and practice. An existing instrument undergoing modifications or being used outside of its target population requires data validity and reliability checks with as much rigor as does a new instrument that is operationalizing a construct. This section will present the validity and reliability tests that were conducted on data resulting from the CSEAS and COES.

Validity

Validity refers to the extent that an instrument measures what it is intended to measure (Barbera & VandenPlas, 2011). Construct and criterion-related validity were the two main types of validity that were assessed in the CSEAS and COES.

a) Construct validity refers to whether a construct has been operationalized accurately. It evaluates whether the instrument is actually measuring the construct it is supposed to measure.

Factor analyses and student interviews were used to evaluate construct validity in the CSEAS and COES.

b) Criterion-related validity evaluates the extent to which an operationalized construct relates to some external criteria (Drost, 2011). Subcategories of criterion-related validity were examined to assess a particular type of validity measure:

(i) Predictive validity evaluates the instrument's ability to predict an external variable based on theory. This was done using correlation analyses which assessed the relationship between the instrument's measures and performance indicators such final exam scores or past performance such as placement test scores.

(ii) Convergent and discriminant validity examine the degree to which an instrument's measures are related or unrelated to other operationalized measures (Barbera & VandenPlas, 2011). Correlation analyses were used to assess the relationships between self-efficacy and anxiety within the CSEAS and self-efficacy and outcome expectations between both surveys. In addition, other measures in the CSEAS were also correlated with self-efficacy factors to test for convergent validity.

Convergent validity was also established qualitatively by verifying whether factors resulting from each survey were similar to the item groups students created during semi-structured interviews.

Reliability

Reliability examines the quality of measurement, particularly the random error in observed data (Barbera & VandenPlas, 2011; Trochim, 2000). Reliability for the CSEAS and COES was assessed using estimates of internal consistency. Cronbach's alpha (α) – mathematically, the average of all possible split-half correlation estimates – was used as the reliability measure for the

CSEAS and COES (Barbera & VandenPlas, 2011; Trochim, 2000). SPSS version 22/23 was used to determine this value for both instruments. Ranging from 0.00 to 1.00, good internal consistency is indicated by alpha values greater than 0.7. However, values higher than 0.9 indicate potentially redundant items in an instrument (Barbera & VandenPlas, 2011). As the subscales (factors) within the CSEAS and COES were measuring different aspects of the same construct, Cronbach's alpha was reported for each subscale as opposed to one alpha value for the CSEAS and COES respectively.

In addition to Cronbach's alpha, item total correlations and square multiple correlations were also examined to assess the reliability of both survey instruments.

Data cleaning

The following measures were implemented to provide the 'purest' data sets for analyses:

a) Given the tendency of students to picket fence in self-report instruments, any student whose survey responses showed zero variance was excluded from analyses.

b) As datasets from different semesters of the same course were combined for analyses, a student retaking the course was included based on their first attempt at the survey because this was their initial, 'pure' response free of any bias.

c) Unless stated otherwise, listwise deletion (from SPSS) was used for all datasets with missing data; cases with multiple responses, missing ID, or a response beyond the intended range were also excluded.

Human Subjects Approval

Research involving data collection from human subjects requires ethical training and approval by the local institutional review board (IRB). The methods utilized, data collected and authorization to disseminate findings were compliant with the IRB protocols at University of

Wisconsin – Milwaukee. Students signed one consent form at the beginning of each semester to designate their participation or lack thereof in data collection procedures for all class-wide surveys administered during that semester. A separate consent form under the same IRB # was used for data obtained using think aloud student interviews. Both forms can be found in **Appendices B** and **C**.

CHAPTER 4: DEVELOPMENT AND VALIDATION OF THE CHEMISTRY SELF-EFFICACY AND ANXIETY SURVEY

This chapter describes the development and psychometric testing of data produced by the Chemistry Self-Efficacy and Anxiety survey (CSEAS).

Background and Rationale

Self-efficacy has been a much researched construct ever since Hackett and Betz (1981) proposed that women were severely underrepresented in scientific and technical fields due to their relative lack of preparation in mathematics and their subsequent avoidance of math; while this phenomenon has been attributed to negative attitudes (Fennema & Sherman, 1977) and math anxiety (Richardson & Suinn, 1972), Hackett and Betz explored this investigation by extending Bandura's (1977) self-efficacy theory to the domain of mathematics; they hypothesized that college females have weaker mathematics self-efficacy beliefs than college males and that these beliefs play a key role in the career decision making process, particularly with regards to selecting science-based majors. The first step towards exploring mathematics self-efficacy expectations was to develop a measure of mathematics self-efficacy expectations (Betz & Hackett, 1983). The mathematics self-efficacy scale (MSES) was operationalized to include perceived self-efficacy in three domains: Solving math problems, everyday math tasks and college coursework. The final version of the MSES consisted of 52 items and requested students to indicate their confidence in their ability to "successfully perform the task, solve the problem, or obtain a grade of "B" or better in the college course" (Betz & Hackett, 1983, p332). Students responded using a 10-point rating (0=not confidence at all to 9=complete confidence). The results showed that males displayed stronger self-efficacy expectations than females on most of the items in the MSES, with an emphasis on those related to college coursework. When the tasks consisted of stereotypically feminine activities such as ‘grocery shopping’, the self-efficacy expectations of females were equivalent to those of males. These results brought up the question of whether similar phenomena existed in other domains and if self-efficacy expectations were a potential contributor to these behaviors. Consequently, several researchers began to develop and validate surveys to assess self-efficacy in various academic domains.

The Science Self-Efficacy Questionnaire (SSEQ), consisting of 27 items, was developed to examine high school students’ self-efficacy in science; students responded to the prompt “how much confidence do you have about doing each of the behaviors” on an A-E letter scale with A=quite a lot and E=very little. A pilot test and subsequent EFA revealed four factors, three of which related to self-efficacy in biology, physics and chemistry and the fourth factor related to laboratory self-efficacy (Smist, 1992). When the SSEQ was administered to college students enrolled in general chemistry, the same factors were observed with females displaying significantly lower self-efficacy than males on only the factor related to laboratory self-efficacy.

The LAESE (longitudinal assessment of engineering self-efficacy) instrument was developed and validated to measure the self-efficacy of women studying engineering, in addition to measuring outcome expectations and feelings of inclusion (Marra et al., 2005). The survey resulted in six subscales: Engineering career expectations, engineering self-efficacy I and II, feeling of inclusion, efficacy in coping with difficulties and math outcomes efficacy. In addition, the survey also includes several self-reported measures related to students’ persistence in their degree plans.

Baldwin et al., (1999) developed a college biology self-efficacy instrument to assess students’ self-reported confidence in understanding and utilizing biology in their lives. This survey (Biology Self-efficacy scale), containing 23 Likert-type items, was administered to nonbiology-major students; subsequent EFA resulted in three factors: Methods of biology, which reflected students' perceived confidence in using analytical skills to conduct biological experiments, generalizing skills learned in biology to other biology/science courses and students' confidence in their ability to apply biological concepts to everyday occurrences.

In the chemistry domain, Coll, Dalgety and Salter (2002) developed the Chemistry Attitudes and Experiences Questionnaire (CAEQ) to measure the impact of students' learning experiences on their attitudes towards chemistry and their chemistry self-efficacy. The final version of the CAEQ comprised of three scales – Attitude towards chemistry scale (21 items across all attitudinal subscales), self-efficacy scale (17 items with five items per subscale) and the learning experiences scale (31 items with five items per subscale). These subscales measured attitudes towards chemists, chemistry in society, career interest in chemistry, lecture learning experiences, tutorial learning experiences and such. The self-efficacy component of the CAEQ, for which students indicated how confident they felt about undertaking the specified tasks, was measured using a 7-point semantic differential scale from 'very confident' (7) to 'not confident at all' (1).

Bauer (2008) developed a survey instrument – Attitude toward the Subject of Chemistry Inventory (ASCI) - for measuring student attitudes regarding chemistry. The 20-item survey used a 7-point semantic differential format with students indicating their attitudes toward chemistry on the 7-point scale between two polar opposite adjectives. The survey was administered to students in a general chemistry course; most of these students were in their first year and represented diverse majors, including engineering, sciences, and liberal arts. Students took the survey across an entire week near the end of the semester (Bauer, 2008). EFA results showed the presence of three distinct factors: Interest and Utility, Anxiety and Intellectual Accessibility; some items did not emerge as separate factors and either loaded weakly across the three factors or appeared as a standalone item

(Bauer, 2008). One item was very distinct and called the “fear” item, although it never loaded on the anxiety factor. A second set of items resulted in weak loadings across the three major factors; these items were called the ‘emotional satisfaction’ item set as opposed to a factor (Bauer, 2008).

Uzuntiryaki and Aydin (2009) developed and validated the Chemistry self-efficacy scale for college students (CCSS) to assess college students’ beliefs in their ability to perform essential chemistry tasks. The 21-item survey requested students to indicate their opinion about various statements related to chemistry tasks; the statements were phrased as questions which examined students’ ability in terms of “how well they could” or “to what extent could they”. Students responded on a scale ranging from 1-9 (1=very poorly and 9=very well with interim ratings of ‘poorly’, ‘average’ and ‘well’). EFA resulted in three factors: self-efficacy for cognitive skills, self-efficacy for psychomotor skills and self-efficacy for everyday applications.

While these surveys have been tested for reliability and validity, resulting in ease of item selection, this study aimed to develop or adapt items which would establish an existence of task specificity similar to that in math so as to capture finer changes/fluctuations in self-efficacy that could be tied into specific content areas. The first step to this goal was to capture the greatest changes in self-efficacy in a pre/post manner for the course context so as to replicate literature studies. This necessitated the construction of a meaningful measure of self-efficacy. Thus, the following objectives guided this study:

a) To develop an instrument to assess students’ self-efficacy in chemistry.

b) To establish validity and reliability for the target population of the data resulting from this instrument.

Methodology

This section describes the phases involved in development of the CSEAS. The selection of items, construction of the instrument, testing and participants will also be detailed. The analyses conducted to psychometrically evaluate the resulting data will also be examined.

Development of the CSEAS items – Self-efficacy and anxiety

Conceptualizing the self-efficacy items in the CSEAS involved an adaptive process due to the multitude of global and domain-specific self-efficacy surveys in the literature. In addition to the self-efficacy surveys in chemistry, self-efficacy surveys from other STEM domains were also referred to during the CSEAS development process. As the purpose of the CSEAS was to establish task specificity in chemistry similar to that in math, the MSES, developed by Hackett and Betz (1983) was used to guide the development of similar items and subscales in chemistry. Dalgety and Coll's CAEQ was used for some of the items, despite these items being fairly omnibus in their measurement of self-efficacy. Furthermore, the CSSS, developed and validated recently as a measure of self-efficacy for college chemistry, was used to incorporate several items too. Lab related items were excluded from the CSEAS not only for parsimony but also for utility of this instrument in courses that did not have a laboratory component. Given the highly context- and task-dependent nature of self-efficacy beliefs, the desired level of specificity for the CSEAS had to be selected carefully.

Researchers have recommended generality in efficacy measures if self-efficacy beliefs are being used to explain performance within a generic setting, such as student performance in a chemistry class (particular course). In such situations, there is a tendency for people to make efficacy judgments across the entire range of tasks required by that setting (lab, quizzes, discussion) (Lent & Brown, 2006). Generality is also favored when the breadth of tasks to be considered in the evaluation of efficacy is not fully realized. On the other hand, if the purpose were to predict student performance levels for a particular task, such as a certain type of chemistry problem, a highly specific measure of self-efficacy would be needed to provide the best predictive power. In general, efficacy beliefs offer the most accurate predictions when measures of these beliefs correspond highly with performance criteria (Lent & Brown, 2006).

In this study, the item pool was generated giving consideration to generality, domain and task specificity. Despite the strong emphasis on these three dimensions and their role in distinguishing self-efficacy from other conceptualizations such as self-concept (Pajares, 1996), care had to be taken in adapting statements so that they adequately differentiated self-efficacy from self-concept. While the latter is a person's general perception of self in given fields of functioning, self-efficacy is a person's expectations of what he/she can accomplish in given situations. The evaluation of each of the constructs requires the students to pose different questions about themselves; while self-efficacy beliefs require assessments of confidence and answer the question "can I do this?", self-concept beliefs necessitate reflections of "feeling" or "being" and assess answer the question "how do I feel about myself in a subject specific context?" Thus, the items and prompt to measure self-efficacy using the CSEAS had to be fairly specific for students to immediately pose questions about the confidence in their abilities to perform the given tasks as opposed to making self-appraisals.

In addition to task and domain specificity, self-efficacy judgments are situation-specific (contextual) as well. Context-specificity is crucial in measures of self-efficacy because the accuracy of self-efficacy judgments requires careful consideration of all the affordances and constraints of the task-performing situation (Pajares, 1996). The case for the contextual and mediational role of self-efficacy in human behavior can be examined by exploring the four sources of efficacy beliefs: Mastery experiences, vicarious experiences, verbal persuasions and physiological states. While mastery experiences are the most influential source of these beliefs as they relate to an individual's successful performance, vicarious learning and verbal persuasions are less impactful on efficacy beliefs. On the other hand, physiological states such as anxiety, stress and mood can be influenced by self-efficacy and affect students' performance; studies have shown that mathematics anxiety can often affect students who, ordinarily, do not experience anxiety in other subject domains (Cates & Rhymer, 2003). While the effects of math anxiety are specific to an individual, students with high levels of math anxiety are more inclined to develop negative attitudes towards math and are more likely to avoid taking math courses in college. Given the unfavorable impact of anxiety on self-efficacy and achievement, this study developed a short scale to measure chemistry students' anxiety; while the utility of this scale has been primarily for validation purposes, the relationship between self-efficacy and anxiety might offer added insight into profiles of students who might be at-risk of dropping the course due to the adverse impact of anxiety on their self-efficacy. In addition, anxiety differences and test performance between male and female students could offer empirical support for stereotype threat.

The item pool for the anxiety survey was gathered from two surveys:

a) Revised Mathematics Anxiety Rating Scale (RMARS), developed by Plake and Parker (1982). This survey consisted of 16 items and resulted in two factors – Anxiety related to learning mathematics and anxiety related to mathematics evaluation.
b) Derived Chemistry Anxiety Rating Scale (DCARS), developed by Eddy (2000). This 36-item survey was used to measure anxiety related to learning chemistry, being evaluated in chemistry and handling chemicals.

Items related to anxiety about handling chemicals were not included in the CSEAS. The items were selected based on the extent to which they shared similar contexts or aspects with the self-efficacy statements.

Lastly, when selecting a prompt for this survey, the question of whether self-efficacy and confidence were evaluating the same construct was addressed. While several surveys using 'confidence' to assess self-efficacy have been psychometrically tested as extremely valid and reliable, understanding the differences between these terms was essential to the interpretation of the survey and in the development of related assessment measures. According to Bandura (1997), confidence refers only to the strength of certainty of one's beliefs, and without the need for a positive outcome, such as an individual being completely confident in failure. Despite "confidence" not being synonymous with self-efficacy, when expressed positively, it can be viewed as a component of self-efficacy (Marra & Bogue, 2006).

Structure of the CSEAS

As the structure of the CSEAS varied from the pilot administration (Fall 2012) until it was first administered online (Spring 2013), this section will describe the development by semester and the accompanying changes.

<u>Fall 2012</u>

The structure of the pilot CSEAS survey was adapted from Zaracova et al., (2005) and had 30 Likert-type items modified from existing self-efficacy surveys. Two versions of this survey were administered to evaluate and compare the resulting data and finalize a version for subsequent semesters. The first version required students to assess just their self-efficacy by responding to 30 items using the prompt "How confident are you about:", with the scale ranging from (1 = not confident at all to 5 = totally confident and 6 = not applicable/not sure). The second version required students to assess their self-efficacy and stress for the same 30 items using the prompts "How confident are you about:" for self-efficacy (same scale as above) and a new prompt "How stressful are you about:" for stress on a scale of (1 = not stressful at all to 5 = totally stressful and 6 = not applicable/not sure). These two scales were utilized not only to facilitate psychometric validation – college-related stress has been found to be inversely related to academic performance among traditional undergraduates (Pritchard & Wilson, 2003) – but also to assess students' efficacy beliefs about handling a task they perceived as a threat or a challenge as this would indicate use of coping strategies and persistence at managing the task (Zaracova et al., 2005).

In addition, an anxiety survey was also administered and required students to respond to 12 Likert-type items using the prompt "How anxious do you get when:", with the scale ranging from (1 = not anxious at all to 5 = totally anxious and 6 = not applicable/not sure). In order to measure stress and confidence in the same direction for high stress to correlate with low confidence, the confidence scale was recoded so that raw confidence scores of 1-5 were scored as 5-1 while raw stress scores were scored as rated from 1-5. A rating of 6 on both scales was recoded to zero, which was further changed to a blank (missing value) so as to not impact parametric statistical analyses. The pilot versions of the self-efficacy, self-efficacy and stress and anxiety surveys are shown in **Appendices D, E** and **F** respectively.

<u>Spring 2013</u>

Online administration of the survey was started in Spring 2013; based on analyses of the pilot survey and in an effort to try and gather much more information on an online platform, this survey version excluded the assessment of stress and consisted (in this order) of 30 slightly revised self-efficacy items, 15 items to evaluate anxiety (three of these items were verification items that were added based on student interviews for the pilot version) and a plethora of items which required students to select possible reasons for their self-efficacy and persistence in the course, their likelihood of persisting in a major, information about their current majors, study habits and their sources of help when they struggled in the course (Seymour & Hewitt, 1997; Grunert & Bodner, 2011).

Student interviews – instrument development and implementation

The purpose of this phase of semi-structured student interviews (conducted in Fall 2012) was to aid in instrument development by refining items (from the pilot version) based on student interpretation. The graduate student solicited for participants during the last five minutes of lecture, where the class was informed of the goal of the interviews, compensation for voluntary participation, recording and videotaping of the process and confidentiality guarantees. Compensation was a $20 gift card to the university book store and the interview was scheduled to last 45-60 minutes. 15 students signed up to participate in semi-structured interviews, during which paper copies of the CSEAS (with stress and self-efficacy) and anxiety surveys were presented and students responded to each statement while verbalizing their thoughts about the items and associated contexts.

During interviews, several students required clarification about assessing their self-efficacy in tasks which involved solving a quantitative problem; these students started solving the problem and based their ratings on whether they could successfully solve the problem during the interview. Consequently, these students had to be instructed to assess 'perceived' confidence as opposed to confidence when actually performing the task. While there were revisions made to the self-efficacy items from pilot to online versions, these revisions were minor and were especially pertinent with regards to omnibus items because some students characterized these items as being "too vague". For instance, the item "determining what answer is required from a written description of a problem" was revised to "understanding what a written chemistry problem is asking you to do" because of problems related to context/scope (chemistry vs. other domains) and clarity ("written problems do not always involve answers that can be firmly determined" or "the word determined sounds like it relates to a mathematical problem"). Similarly, in a different item, the term "formula" was replaced with "equation" due to students' familiarity with the latter term. Some of the items were revised because of the idea that students' self-efficacy would plateau or stay unchanged due to the overly simplistic nature of the item; for example, "converting the temperature in your home from degrees Celsius to Kelvin" was changed to "converting the temperature from Fahrenheit to Kelvin". One of the items that posed some issues even after revisions was "learning chemistry in this chemistry course (if there were no exams to take)". Most students stated that "without exams, they would not know how well they had learned material" and were inclined towards the neutral option on the survey. When this item was revised to "learning chemistry (if all exams were take-home exams)", some students gravitated towards indicating a confidence level due to the presence of an evaluative component, albeit a take-home exam, that would allow them to use what they had learned.

With regards to the stress component in the CSEAS, research supports the assumption that awareness of a negative stereotype increases situational anxiety, stress and fear of evaluation (Steele & Aronson, 1995). For instance, a negative stereotype exists about female students being inferior in math relative to male students; if female students were to be evaluated or judged in terms of a negative stereotype, they would be likely to perform worse in a domain in which their subgroup experiences negative stereotypes. Given that negative effects of stereotype threat on performance are mediated by physiological states such as stress, female students would be doubly threatened about not only their perceived confidence on math items, but also being reduced to a negative stereotype targeting their group in the chemistry domain (Steele & Aronson, 1995). There were no apparent of indications of stereotype threat during the interviews; some female students who exhibited an innate stress or low self-efficacy for certain items or throughout the interview process were those who were non-traditional students or had not taken a chemistry class in several years. From a psychometric perspective, while the general prediction was that stress and confidence would demonstrate an inverse relationship, this was not a consistent and clear relationship. Students' responses to stress and confidence for a statement were highly dependent on the appraisal of the task and whether it was deemed a challenge or a threat (Chemers et al., 2001). For instance, while the students were not very confident about explaining why salt melts ice, there were not stressful at all about the task due to the minimal threat it posed from an assessment or evaluative standpoint. Another item that displayed similar issues was regarding students' confidence about receiving the grade they desired in the course. While some students were confident of their study habits and of the grade that would be received, they were highly stressed because "sometimes hard work did not necessarily translate to an expected grade". Thus, subsequent to pilot testing, the CSEAS administered online excluded the 'stress' scale in order to avoid ambiguities about the purported relationship between stress and self-efficacy.

For the anxiety survey, some students had a different interpretation of 'anxious' and viewed it as excitement or anticipation instead of making a negative association. This required the original anxiety survey to be amended to included three verification items. These are items that are inclined to elicit either an absolutely positive or absolutely negative response with little to no room for interpretational ambiguities. Scale for the anxiety survey was 1=not anxious at all to 5=totally anxious and 6= not applicable/not sure. The verification items added to the anxiety survey were students' anxiety related to:

a) Getting extra credit for attending your chemistry lecture

b) Cramming the night before your chemistry exam

c) Not knowing the material on your chemistry exam

Thus, if students indicated low anxiety levels for items (b) or (c) or high anxiety levels for item (a), these students were perhaps interpreting anxious differently and were subsequently excluded from analyses. The final version of the online CSEAS is included in **Appendix G**.

The purpose of this second phase of semi-structured student interviews was to aid in psychometric testing by establishing another layer of validity. Additionally, these interviews also aided in understanding context sensitivity by examining the degree to which context/perception of a student affected their perceived confidence.

The graduate student solicited for participants during the last five minutes of lecture, where the class was informed of the goal of the interviews, compensation for voluntary participation and confidentiality guarantees. Compensation was a $10 gift card to the university book store and the interviews were scheduled to last 30-45 minutes.

During the interview, the researcher demonstrated the task at hand by using an example unrelated to chemistry. Notecards with culinary references and items were placed on the table as shown in **Figure 4.1**. The researcher then demonstrated the task by sorting the cards into appropriate categories as shown in **Figure 4.2**.

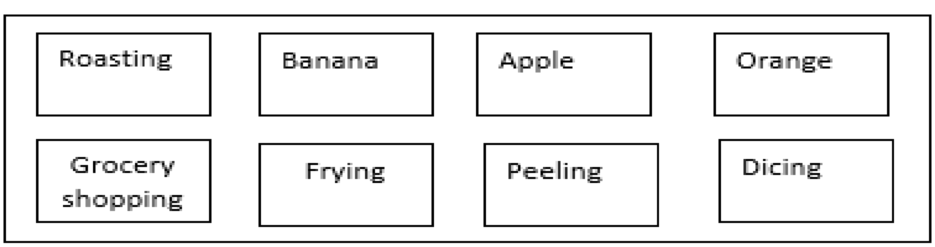


Figure 4.1. Example of notecards used for sorting task during interviews

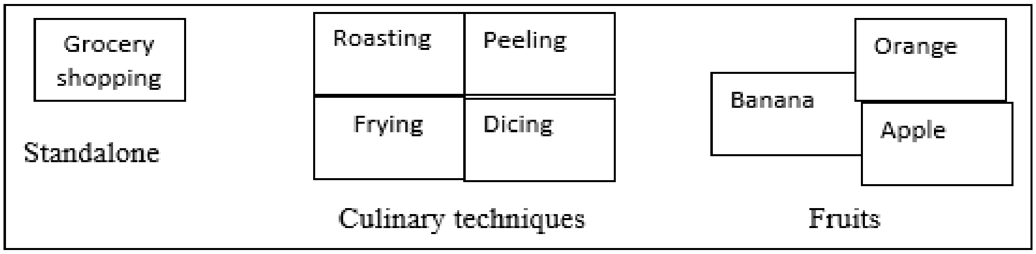


Figure 4.2. Example of notecards sorted into meaningful groups

Following this demonstration, students were given notecards, each with the name and number of a self-efficacy item from the CSEAS as shown in **Figure 4.3**. Students were asked to group these items and offer a category name.

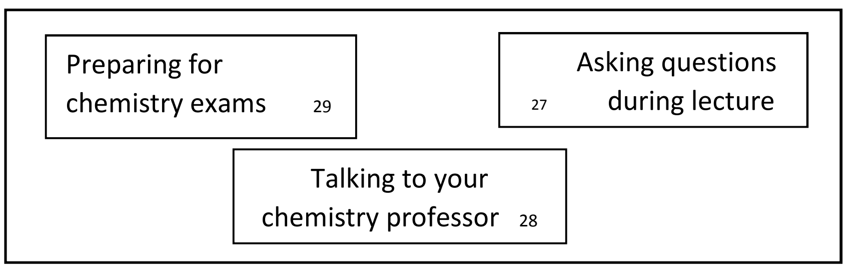


Figure 4.3. Example of notecards with actual survey items and numbers for sorting task during interviews

Students listed their categories and constituting items, following which the researcher compared their groupings to those obtained from factor analyses. In addition, groupings created by interviewees were also compared to each other to evaluate potential similarities in context association, if any.

Data collection and participants

A summary of the measures collected, time point and purpose of testing is shown in **Table 4.1.**

Table 4.1. Summary of testing purpose, timeline and data collected - CSEAS

Testing	Purpose of testing	Time point	Data collected
1	Pilot survey administration (paper)	Fall 2012	a) Survey with standalone self-efficacy prompt b) Survey with self-efficacy and stress c) Anxiety survey
2	Interviews - pilot instrument development and psychometric testing	Fall 2012 and Fall 2013	a) Stress-self-efficacy survey and anxiety survey b) Notecards with item names (for grouping)
3	Main survey administration (Qualtrics)	Since spring 2013	a) Self-efficacy, anxiety, contextual measures and intended major

The first attempt at measuring self-efficacy was carried out in Spring 2012 using a semantic differential instrument (Bauer, 2008). As data was collected by a different researcher, the demographic details of participants have been excluded from this section. However, a summary of the results and excerpts from interviews conducted with two students will be presented in the appendix as part of the results and discussion for this preliminary data.

The CSEAS has been in administration since Fall 2012; the pilot study using CSEAS on paper was conducted in Fall 2012 while online data collection using the final version of the CSEAS has been occurring since Spring 2013.

In fall 2012, the two versions of the CSEAS paper survey were tested on two different lecture sections of GC I to compare resulting factor structures and decide on a version for administration in subsequent semesters. Each section of students was given a version of the self-efficacy survey and the anxiety survey by the course instructor in lecture during the first week of class. For courses other than GCI, surveys were distributed and collected by teaching assistants during their discussions. Each discussion section received either version of the self-efficacy survey and the anxiety survey.

For GC I, course instructor explained the purpose and importance of the surveys during lecture, in addition to extra credit incentives that would be offered upon completion of the surveys.

Students who returned incomplete surveys or did not return their surveys within the first two weeks since start of classes were not included in data analyses as there was a chance that these students had been sufficiently exposed to the course material and to the instructor for their responses to be biased. The post surveys (two versions) for both lecture sections of GC I were distributed a week or two before the start of final exam week. Surveys were distributed and collected in the same lecture. The paper surveys (self-efficacy and anxiety) typically took 10-15 minutes to complete.

Starting in spring 2013, the CSEAS was distributed online using a link generated by Qualtrics, the platform that housed the survey. As this link was common to another survey (assessing students' knowledge of scale) that was already being administered to GC I students, the CSEAS survey was attached to this survey. Thus, students responded to the scale survey followed by the CSEAS. This sequential administration of two surveys was followed for GC I only. All other courses were sent links that only administered the CSEAS. Links for the pre-surveys were sent out to each course instructor along with a brief greeting to the students, explaining extra credit incentives and the duration for which the link would be open. As these incentives were specific to each course, instructors modified the incentives as they saw fit before sending out the link to their students. Links were sent out a day or two before classes would start and were active for one week. While there was no official "deactivation" of the survey, students who submitted their surveys past the closing deadline (as indicated by timestamps associated with each submission) were not included in analyses. The CSEAS portion of the survey typically took about 20 - 30 minutes to complete although it was possible to complete the survey in about 6 minutes if responses were clicked blindly. Post-survey links (following the same protocols as described above) were sent out a week or two before the start of final exams week.

The studies described in this chapter were conducted at a large, urban, research intensive public university in the Midwestern United States. Surveys were administered to students enrolled in preparatory chemistry, GC I, GC II and general chemistry for engineers; the descriptions of these courses are given in chapter 3. **Tables 4.2** and **4.3** show the participants for the pilot and online administrations of the CSEAS. **Table 4.2** shows participants for the version of the survey that incorporated the standalone self-efficacy scale (without the stress scale) as this standalone version was going to be in use for subsequent semesters.

Table 4.2. Participants (by course) for pilot administration of paper version of CSEAS – Fall 2012

Fall 2012 (pilot)	Prep. Chem	Gen. Chem. I	Gen. Chem. II	Gen. Chem. for engineers
Pre (N) - SE only	84	155	58	39
Pre (N) - Anxiety	156	150	55	65
Post (N) - SE only	65	85	78	50
Post (N) - Anxiety	67	100	110	43

Table 4.3. Participants (by course) for administration of online version of CSEAS – Spring 2013

Spring 2013	Prep. Chem	Gen. Chem. I	Gen. Chem. II	Gen. Chem. for engineers
Pre (N)	166	181	102	36
Post (N)	80	114	60	N/A

Data analyses

Data were cleaned as described in chapter 3. For the pilot version which had both self-efficacy and stress assessments, the confidence (self-efficacy) scale was recoded so that raw confidence scores of 1-5 were scored as 5-1 while raw stress scores were scored as rated from 1-5. This was done in order to measure stress and confidence in the same direction so that a high stress score would correlate with low confidence scores. A rating of 6 on both scales was recoded to zero, which was further changed to a blank (missing value) so as to not impact parametric statistical analyses. Although the online version of the survey excluded the 'stress' prompt, recoding the self-efficacy scale as described above was continued to stay consistent.

An additional piece of information that had to be noted since the online administration of the survey was the timestamp associated with each student's response. As there were timestamps that were either too low (possible lack of variance in student responses) or two high (if students had the survey open, moved on to other tasks and returned to complete it), a plot of frequency vs. timestamp, as shown in **Figure 4.4** was observed for normality. These plots, observed each semester, show an average time of an hour for completing both surveys that are part of the link with some students who fall into timestamp ranges well below and beyond the mean. While students on the lower end often coincided with those who lacked any variance in their responses and were excluded on this basis as part of data cleaning, students on the upper end displayed considerable variance in their ratings, and had typed out text for some of their open response items. Therefore, for the purposes of this study, only students who had zero variance in their data were excluded from any analyses. While timestamps continue to be noted, excluding students on the basis of short completion times because they may not have responded to the survey thoroughly was not the approach that was taken here.

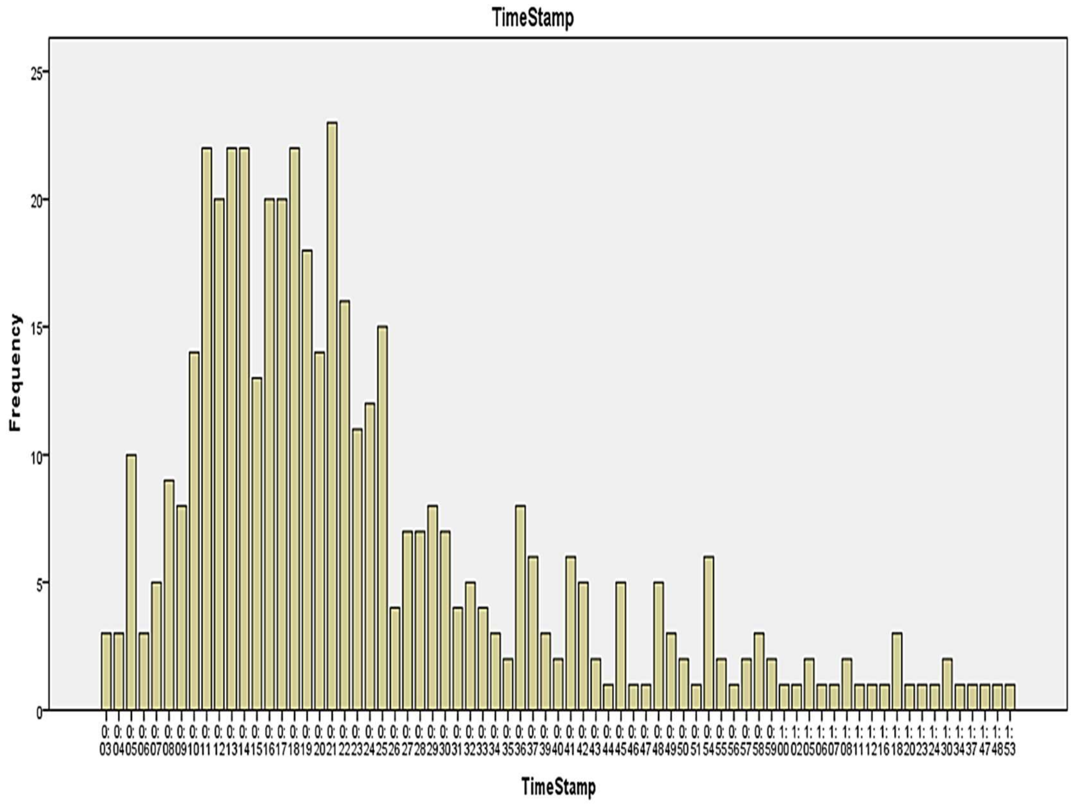


Figure 4.4. Plot of frequency of students against online survey completion time for Fall 2014

Descriptive statistics were obtained for all items in the CSEAS for assessments of univariate normality, skew, kurtosis and missing data.

<u>For GC I alone</u>

EFA was conducted in an effort to replicate the factor structure resulting from data collected at another institution.

<u>For data from GC I and GC II</u>

Factor analysis (EFA and CFA) was conducted to determine the most robust and meaningful factor structure. An average score was calculated for each subscale ("factor") based on the student responses to statements (items) and the high degree of relatedness of the items that constituted the subscale. As the overall goal of this project was to obtain consistently stable measures of self-efficacy and outcome expectations at different time points during a longitudinal model, EFA was meant to be conducted using datasets from courses that frame the different time points in this model and comprise of fairly homogeneous groups of students in terms of ability levels – post-GC I, pre-GC II and a combined dataset respectively. However, as GC II had different instructors for a semester or two, there were lapses in data collection due to delayed distribution of survey links, which resulted in exclusion of data from these semesters. As a result, while sample sizes were adequate for conducting EFA on post-GC I datasets, they were not large enough for EFA to be conducted on pre-GC II datasets by themselves. Therefore, EFA were conducted on two combined datasets from Fall 2012 and Spring 2013. These results were compared to an EFA that was run using post-GC I data alone.

Comparative statistics were obtained (using GC I) as described in chapter 3. For independent sample t-tests, high vs. low performing groups (on final exam and in the course) were created based on z-scores for the raw data. Students with z-scores > 0 were categorized as the high-performing group while z-scores < 0 were the low-performing group. In the CSEAS, a low average score on a self-efficacy subscale implied high self-efficacy (confidence) while a low average score on a stress or anxiety subscale implied low stress or anxiety.

Reliability and validity were established using the measures detailed in chapter 3.

<u>For data from preparatory chemistry and general chemistry for engineers</u>

Cluster analyses were used to group the responses from students in these courses. While these courses do not play an integral role in the development of a longitudinal model, they serve as two key courses that pave the way for students to be primed for enrollment in general chemistry or in their respective engineering fields. Thus, the analysis conducted here is the first step to establish affective and cognitive meaning in two courses comprising of highly heterogeneous groups of students.

Results and discussion

Semantic differential

. The quantitative and qualitative results of this analysis, using data collected prior to Spring 2012, are summarized in **Appendix H**.

Descriptive statistics

The demographic statistics for combined datasets from Fall 2012, Spring 2013 and post-GC I from Fall 2014 are provided in **Table 4.4**.

Table 4.4. Demographic characteristics of a) combined datasets from Fall 2012, Spring 2013 and post-GC I from Fall 2014

Post-GCI + Pre-GC II - Fall'12			Post-GCI + Pre-GC II - Spring'13			Post-GCI - Fall'14		
Variable	N	%	*Variable*	N	%	*Variable*	N	%
Gender			Gender			Gender		
Male	56	39.4	Male	84	39.6	Male	49	45.0
Female	86	60.6	Female	128	60.4	Female	60	55.0
Acad.Level			Acad.Level			Acad.Level		
Freshman	13	9.2	Freshman	34	16.0	Freshman	21	19.3
Soph.	64	45.1	Soph.	82	38.7	Soph.	52	47.7
Junior	29	20.4	Junior	44	20.8	Junior	20	18.3
Senior	36	25.4	Senior	52	24.5	Senior	15	13.8
ACT scores	Average		ACT scores	Average		ACT scores	Average	
Composite	24.29		Composite	23.92		Composite	23.60	
Math	24.18		Math	24.09		Math	22.99	
Sci.	24.39		Sci.	23.97		Sci.	23.75	

Among the intended majors, on average, about 30% of students were biology, microbiology and biochemistry majors, while close to 20% were undecided followed by 12% of students with various majors in liberal arts and the remaining percentage allocated to majors in miscellaneous fields such as business, engineering and education. Although the data show ACT scores for combined datasets and one post-GC I dataset, when t-test results were examined for differences in ACT scores between students in post-GC I and pre-GC II, no significant differences were observed for ACT scores in either component or overall; consequently, a combined dataset of students from post-GC I and pre-GC II was used for factor analyses. The fall 2012 combined dataset had the highest percentage of missing data (2.8%) for items 9, 18 and 21 while the spring 2013 combined dataset had 2.4% missing data for items 10, 18 and 20 and 2.9% for items 28 and 25. While some individual items displayed skewness and kurtosis above recommended values, the resulting subscales or factors for the final model had values within range.

Factor analysis – Exploratory

Table 4.5 shows the factor structure of two combined datasets from Fall 2012, Spring 2013 and a post-GC I dataset from Fall 2014. These structures shown items in sequence and how these items tracked across the structures; decisions were made about items that needed to be excluded from analyses. The correlation matrix and item characteristics are only shown for the dataset that lead into the final factor structure.

Table 4.5. Component matrices: Factor structures of datasets from Fall 2012, Spring 2013 and Fall 2014

N = 109

Item	Seven factors - Post-GC I (Fall 2014) ONLINE						
	1	2	3	4	5	6	7
1	.523	.679					
2		.675					
3		.628					
4		.726					
5				.625			
6		.702					
7				.548			
8		.536					
9						.824	
10					.584		
11			.715				
12			.832				
13			.811				
14	.525						
15			.520				
16				.699			
17				.784			
18							
19					.587		
20	.786						
21	.546						.534
22	.668						
23	.883						
24	.851						
25	.680						
26	.822						
27	.707						
28	.645						
29	.788						
30	.672						

N = 127

Item	Six factors - Post-GC I + Pre-GC II (Fall 2012) PAPER							
	1	2	3	4	5	6	7	8
1		.600						
2		.674						
3		.688						
4			.639					
5			.709					
6			.534					
7			.619	.584				
8								
9		.642						
10							.618	
11							.816	
12								
13						.683		
14					.718			
15					.556			
16				.559				
17				.634				
18				.756				
19								.620
20	.721							
21								
22		.588						
23	.769							
24	.584							
25								
26	.795							
27	.694							
28	.672							
29	.799							
30						.574		

N = 182

Item	Six factors - Post-GC I + Pre-GC II (Spring 2013) ONLINE					
	1	2	3	4	5	6
1	.530					
2	.562					
3						
4		.618				
5		.754				
6		.532				
7						
8		.589				
9						
10						
11			.787			
12			.764			
13		.538				
14						
15		.518		.531		
16				.796		
17				.787		
18						
19						
20	.732					
21					.736	
22	.748					
23	.838					
24	.826					
25	.587					
26	.792					
27						
28					.776	
29	.730					
30	.629					

These structures were extracted using the eigenvalue > 1 condition. Each combined dataset was obtained from semesters when the mode of survey administration was different. The dataset from just post-GC I shows one factor (items 23-14) that comprises much of the assessment and evaluation items. When an EFA was carried out on just this factor, it separated into two factors – one involving interpersonal tasks (items 21,27 and 28) while the other comprised of tasks related to chemistry learning, assessment and evaluation. This dataset also resulted in a factor structure where task specificity was not particularly prevalent. This was not an unusual observation with pre vs. post factor structures in general. Task specificity was more distinct in pre survey results as opposed to post results, where context association perhaps seemed less important than the act of problem solving thereby resulting in mass groupings of items showing little to no contextual delineation.

According to Schunk and Pajares (2001), a key distinction exists between self-efficacy for performance vs. self-efficacy for learning. When there is familiarity with the rigors of an activity or task, students are likely to draw upon self-efficacy beliefs related to prior experiences with similar tasks – these beliefs are called self-efficacy for performance as they are associated directly with an intended performance goal. On the other hand, students might not aware of the skills they require when confronted with unfamiliar tasks. In these situations, students infer their self-efficacy beliefs from past achievements in situations perceived as similar to the new one – these beliefs are called self-efficacy for learning as they are inferred assessments made about one's capability to learn the necessary skills for successful completion of the task (Schunk, 1996; Zimmerman, Bandura & Martinez-Pons, 1992). This could explain why some degree of task specificity might be observed in pre-factor structures in comparison to post-structures.

While combining a pre-GC II dataset to a post-GC I set did assist in resolving factor 1 into smaller groups, it also revealed several problematic items that either did not load at all or was an item all by itself. Some of these items were common to both combined datasets although the combined dataset resulting from the paper survey produced groups that were more meaningful substantively, especially when the problem items were excluded.

As task specificity in factor structure was one of the goals of developing this instrument and subsequently developing a shortened version of this instrument, the factor structure resulting from the Fall 2012 combined dataset was used as the starting point; items were excluded and resulting factor structures were compared to decide on the most meaningful structure. The item means, standard deviations and inter-item correlation matrix for this dataset are presented in **Table 4.6.** On a 6-point scale, where 1=not confident at all and 5=totally confident and ratings of 6=not applicable/not sure were treated as zeros (and subsequently changed to blanks), means ranged from 1.44 to 3.23. The correlation matrix does not show any correlations exceeding r=.70, thus indicating no problems with multicollinearity. Bartlett's test of sphericity was significant (χ^2 = 1588.9, p = .000), which indicated that the correlation matrix was not an identity matrix. The KMO statistic (.80) was between 0.5 and 1 and categorized as 'fair,' indicating the matrix was appropriate for factor analysis. The determinant of this matrix was 1.02×10^{-6}.

Table 4.6. Correlation matrix, means and standard deviations for the chemistry self-efficacy component of the CSEAS. Complete data from 127 students was used for EFA.

Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	Mean	SD
1	1	0.536	0.394	0.243	0.005	0.219	0.078	0.026	0.209	0.122	0.204	0.355	0.104	0.222	0.255	0.198	0.27	0.107	0.118	0.277	0.122	0.406	0.392	0.365	0.269	0.276	0.255	0.123	0.35	0.141	2.08	0.783
2	0.536	1	0.504	0.357	0.085	0.302	0.135	0.175	0.29	0.166	0.014	0.214	0.189	0.32	0.308	0.26	0.199	0.053	0.121	0.324	0.254	0.475	0.368	0.432	0.278	0.388	0.287	0.206	0.261	0.193	2.09	0.702
3	0.394	0.504	1	0.367	0.063	0.301	-0.062	0.065	0.28	0.189	0.196	0.315	0.131	0.32	0.356	0.317	0.198	0.095	0.127	0.153	0.074	0.313	0.231	0.362	0.365	0.235	0.188	0.075	0.101	0.216	1.81	0.742
4	0.243	0.357	0.367	1	0.343	0.44	0.296	0.29	0.203	0.364	0.114	0.199	0.147	0.316	0.341	0.275	0.237	0.065	0.122	0.132	0.164	0.36	0.247	0.268	0.364	0.215	0.056	0.021	0.087	0.161	1.74	0.693
5	0.005	0.085	0.063	0.343	1	0.215	0.336	0.211	0.211	0.269	0.107	0.128	0.108	0.128	0.225	0.1	0.217	0.027	0.157	0.029	0.198	0.045	0.094	0.03	0.11	0.072	0.029	-0.001	-0.088	0.061	1.96	0.858
6	0.219	0.302	0.301	0.44	0.215	1	0.419	0.137	0.225	0.249	0.261	0.147	0.236	0.282	0.204	0.253	0.278	0.15	0.097	0.107	0.127	0.32	0.128	0.209	0.168	0.162	0.159	0.062	0.019	0.18	1.54	0.676
7	0.078	0.135	-0.062	0.296	0.336	0.419	1	0.368	0.123	0.169	0.151	0.132	0.193	0.179	0.2	0.231	0.358	0.379	0.088	0.209	0.123	0.137	0.14	0.071	0.132	0.193	0.161	0.095	0.082	0.058	2.2	0.909
8	0.026	0.175	0.065	0.29	0.211	0.137	0.368	1	0.086	0.03	0.067	0.023	0.102	0.024	0.066	0.202	0.103	0.186	0.152	0.192	0.156	0.235	0.158	0.161	0.219	0.167	0.131	0.133	0.115	0.055	1.84	1.019
9	0.209	0.29	0.28	0.203	0.211	0.225	0.123	0.086	1	0.027	0.162	0.106	0.172	0.092	0.214	0.208	0.177	0.087	0.003	-0.005	0.075	0.276	0.046	0.151	0.156	0.04	0.048	0.002	-0.034	0.085	1.44	0.709
10	0.122	0.166	0.189	0.364	0.269	0.249	0.169	0.03	0.027	1	0.343	0.302	0.139	0.124	0.388	0.236	0.182	0.138	0.167	0.2	0.231	0.23	0.327	0.33	0.151	0.321	0.226	0.176	0.144	0.327	1.62	0.899
11	0.204	0.014	0.196	0.114	0.107	0.261	0.151	0.067	0.162	0.343	1	0.339	0.16	0.15	0.13	0.28	0.151	0.296	0.308	0.022	-0.068	0.148	0.157	0.2	0.059	0.039	0.194	0.068	0.02	0.241	1.63	0.88
12	0.355	0.214	0.315	0.199	0.128	0.147	0.132	0.023	0.106	0.302	0.339	1	0.256	0.394	0.302	0.374	0.301	0.28	0.218	0.257	0.2	0.344	0.237	0.278	0.179	0.175	0.121	0.109	0.183	0.234	1.69	0.972
13	0.104	0.189	0.131	0.147	0.108	0.236	0.193	0.102	0.172	0.139	0.16	0.256	1	0.288	0.423	0.304	0.316	0.195	0.197	0.15	0.263	0.136	0.167	0.078	0.138	0.145	0.31	0.226	0.095	0.346	1.6	0.681
14	0.222	0.32	0.32	0.316	0.128	0.282	0.179	0.024	0.092	0.124	0.15	0.394	0.288	1	0.412	0.266	0.223	0.097	0.195	0.262	0.074	0.21	0.256	0.281	0.146	0.246	0.058	0.05	0.146	0.222	2.17	1.067
15	0.255	0.308	0.356	0.341	0.225	0.204	0.2	0.066	0.214	0.388	0.13	0.302	0.423	0.412	1	0.497	0.383	0.196	0.238	0.276	0.105	0.185	0.219	0.171	0.201	0.188	0.184	0.211	0.141	0.373	1.63	0.664
16	0.198	0.26	0.317	0.275	0.1	0.253	0.231	0.202	0.208	0.236	0.28	0.374	0.304	0.266	0.497	1	0.566	0.387	0.212	0.323	0.092	0.369	0.366	0.359	0.286	0.256	0.261	0.152	0.204	0.328	2.46	1.045
17	0.27	0.199	0.198	0.237	0.217	0.278	0.358	0.103	0.177	0.182	0.151	0.301	0.316	0.223	0.383	0.566	1	0.435	0.212	0.333	0.165	0.246	0.29	0.233	0.219	0.265	0.294	0.143	0.22	0.304	2.8	0.979
18	0.107	0.053	0.095	0.065	0.027	0.15	0.379	0.186	0.087	0.138	0.296	0.28	0.195	0.097	0.196	0.387	0.435	1	0.125	0.22	0.08	0.142	0.204	0.148	0.215	0.148	0.191	0.134	0.152	0.234	2.34	0.91
19	0.118	0.121	0.127	0.122	0.157	0.097	0.088	0.152	0.003	0.167	0.308	0.218	0.197	0.195	0.238	0.212	0.212	0.125	1	0.264	0.091	0.153	0.153	0.089	0.21	0.114	0.179	0.136	0.063	0.302	1.98	1.05
20	0.277	0.324	0.153	0.132	0.029	0.107	0.209	0.192	-0.005	0.2	0.022	0.257	0.15	0.262	0.276	0.323	0.333	0.22	0.264	1	0.328	0.416	0.655	0.502	0.431	0.564	0.471	0.415	0.569	0.176	2.53	1.16
21	0.122	0.254	0.074	0.164	0.198	0.127	0.123	0.156	0.075	0.231	-0.068	0.2	0.263	0.074	0.105	0.092	0.165	0.08	0.091	0.328	1	0.313	0.31	0.255	0.34	0.405	0.25	0.547	0.303	0.155	3.23	1.223
22	0.406	0.475	0.313	0.36	0.045	0.32	0.137	0.235	0.276	0.23	0.148	0.344	0.136	0.21	0.185	0.369	0.246	0.142	0.153	0.416	0.313	1	0.384	0.709	0.436	0.384	0.238	0.242	0.285	0.124	2.02	1.042
23	0.392	0.368	0.231	0.247	0.094	0.128	0.14	0.158	0.046	0.327	0.157	0.237	0.167	0.256	0.219	0.366	0.29	0.204	0.153	0.655	0.31	0.384	1	0.623	0.439	0.604	0.574	0.404	0.596	0.228	2.88	1.117
24	0.365	0.432	0.362	0.268	0.03	0.209	0.071	0.161	0.151	0.33	0.2	0.278	0.078	0.281	0.171	0.359	0.233	0.148	0.089	0.502	0.255	0.709	0.623	1	0.494	0.54	0.333	0.269	0.422	0.09	2.54	1.153
25	0.269	0.278	0.365	0.364	0.11	0.168	0.132	0.219	0.156	0.151	0.059	0.179	0.138	0.146	0.201	0.286	0.219	0.215	0.21	0.431	0.34	0.436	0.439	0.494	1	0.477	0.341	0.348	0.258	0.217	2.43	1.27
26	0.276	0.388	0.235	0.215	0.072	0.162	0.193	0.167	0.04	0.321	0.039	0.175	0.145	0.246	0.188	0.256	0.265	0.148	0.114	0.564	0.405	0.384	0.604	0.54	0.477	1	0.55	0.47	0.604	0.38	2.63	1.053
27	0.255	0.287	0.188	0.056	0.029	0.159	0.161	0.131	0.048	0.226	0.194	0.121	0.31	0.058	0.184	0.261	0.294	0.191	0.179	0.471	0.25	0.238	0.574	0.333	0.341	0.55	1	0.628	0.528	0.381	2.72	1.147
28	0.123	0.206	0.075	0.021	-0.001	0.062	0.095	0.133	0.002	0.176	0.068	0.109	0.226	0.05	0.211	0.152	0.143	0.134	0.136	0.415	0.547	0.242	0.404	0.269	0.348	0.47	0.628	1	0.481	0.226	2.76	1.185
29	0.35	0.261	0.101	0.087	-0.088	0.019	0.082	0.115	-0.034	0.144	0.02	0.183	0.095	0.146	0.141	0.204	0.22	0.152	0.063	0.569	0.303	0.285	0.596	0.422	0.258	0.604	0.528	0.481	1	0.345	3.2	1.157
30	0.141	0.193	0.216	0.161	0.061	0.18	0.058	0.055	0.085	0.327	0.241	0.234	0.346	0.222	0.373	0.328	0.304	0.234	0.302	0.176	0.155	0.124	0.228	0.09	0.217	0.38	0.381	0.226	0.345	1	2.17	0.889

Item-to-total correlations (ITCs) showed low values for some items (items 5 and 9=.22, item 8 = .28); however, no negative ITCs or extremely low square multiple correlations were observed. Under the conditions of PCA as the extraction method, varimax rotation and the default eigenvalue > 1 criterion, an eight-factor solution was obtained. Item 19 loaded by itself, and there were several factors that only comprised of two items. Subsequently, the number of factors to be extracted was entered manually, leaving other conditions unchanged. Factor solutions ranging from four to seven factors were extracted. The six-factor solution displayed a structure in which most subscales containing at least three items per factor and one subscale containing two items. This structure was then examined for problem items that elicited exclusion from further analyses. In the six-factor solution, items 24 and items 7 showed cross-loadings, while no factor loaded items 9,14,6,12 and 19. The scree plot, shown in **Figure 4.5** justified retention of four or five factors.

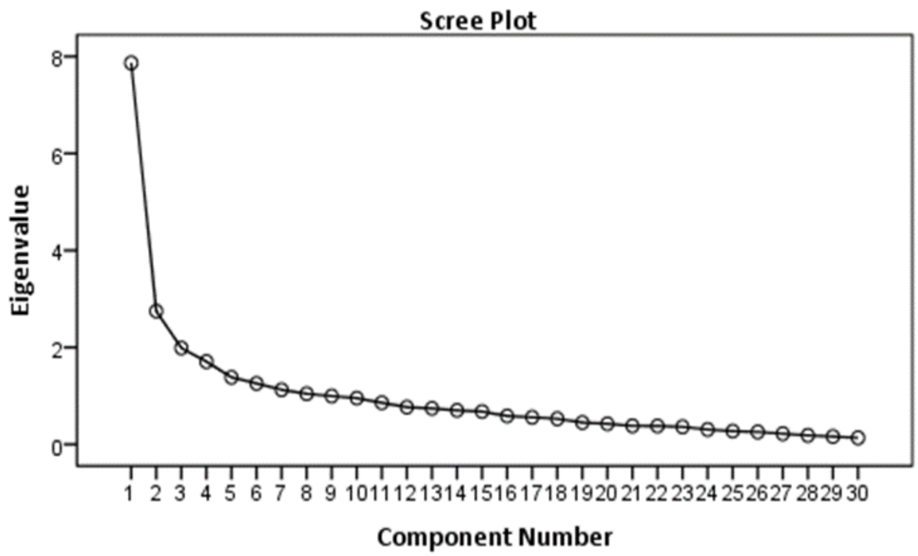


Figure 4.5. Scree plot showing eigenvalues for retaining a four- or five-factor solution

Additionally, parallel analysis results shown in **Table 4.7** recommended retention of four factors.

Table 4.7. Parallel analysis results showing eigenvalues for actual and random ordered data – CSEAS

Eigenvalue #	Actual data eigenvalue	Random order eigenvalue	Percentile
1	7.86757	2.04523	2.18536
2	2.74764	1.88852	1.99689
3	1.98788	1.76786	1.86331
4	1.70715	1.66789	1.74141
5	1.38210	1.57639	1.65083
6	1.25600	1.49498	1.56281

Using the six-factor solution as the starting point and comparing it to other extracted solutions items 9,14,6,19, and 7 were excluded for not loading on any factor or displaying cross loadings consistently. Items 15 and 30 excluded on the basis consistently moving around and being part of factors that were otherwise meaningful. Although item 24 showed a cross loading in the six-factor solution, it did not cross load when some of the problem items were removed and reanalyzed sequentially. A four-factor solution, after item exclusions, was the most conceptually interpretable factor structure. Item 4 had a cross loaded with factors 2 and 4. However, as the difference in loadings was considerable, this item was deemed a part of factor 4. The total variance explained by the factors was 54.3%. Variance explained by each factor was as follows: factor 1 = 19.3%, factor 2 = 15.4%, factor 3 = 11.7% and factor 4 = 8.0%. The factor loadings from the rotated component matrix are shown in **Table 4.8**. Factor names and items in each factor are shown in the table. This factor structure was tested using a different data set for CFA; the final model for the CSEAS is represented in the path diagram. This model accounts for any modifications from CFA to improve model fit.

Subscales were named based on the tasks described by the items. Tasks which involved recalling information or cognitive memory thinking such as trends in the periodic table or using the periodic table to identify elements that are gases were called “low order / recall tasks”. Similarly, those tasks that required divergent thinking where several reasonable answers were possible were categorized as “higher order tasks” (Pavelich, 1982). Tasks related to assessment and evaluation (taking exams, receiving grades and learning chemistry) were named as such and those subscales that consisted of simple or slightly complex problems involving application of chemistry to everyday tasks were labeled using similar terms.

Table 4.8. Rotated component matrix for four-factor solution: CSEAS (post-GC II + pre-GCI; N=127)

Factor & Item	Item	Four factors - Pre GC II + Post GC I			
		1	2	3	4
Factor 1	**Self efficacy related to assessment, evaluation and interpersonal tasks**				
28	Talking to your chemistry professor	.779			
29	Receiving the grade you desire in this course	.752			
26	Preparing for chemistry exams	.737			
27	Understanding your chemistry professor	.729			
20	Doing well on chemistry course exams, given you exert enough effort	.705			
23	Taking an exam or quiz in your chemistry course	.685			
21	Asking questions during lecture	.556			
Factor 2	**Self efficacy related to applying problem solving strategies**				
3	Determining appropriate units for a numerical result		.694		
2	Choosing an appropriate equation to solve a chemistry problem		.693		
1	Understanding what a written chemistry problem is asking you to do		.681		
24	Taking a chemistry exam or quiz where considerable math is involved		.665		
22	Learning material in chemistry courses where considerable math is involved		.650		
25	Signing up for more chemistry courses in the future (regardless of the outcome of this course or the requirements for your major)				
Factor 3	**Self efficacy related to higher order and applying chemistry to everyday tasks**				
18	Writing a summary of the main points of a television documentary that deals with some aspect of chemistry			.689	
16	Explaining why addition of salt melts ice			.671	
17	Using chemistry to propose a solution that keeps cooking water from boiling over			.661	
11	Converting your speedometer reading from mph to yards/second (1 mile = 1760 yards)			.614	
12	Calculating the density of lemonade (made by adding 50g of lemons to 500mL of water)			.552	
13	Identifying the type of change (physical vs. chemical) when milk gets sour			.520	
Factor 4	**Self efficacy related to low order / recall tasks**				
5	Describing trends in the periodic table (atomic size, electronegativity)				.698
4	Reading and writing a chemical formula		.514		.594
8	Identifying elements that are gases at room temperature (from the periodic table)				.568

Although the anxiety survey was utilized primarily for psychometric validation of the self-efficacy subscales, a sound factor structure was required for this purpose as well. EFA was conducted on data resulting from the anxiety surveys (after excluding students who responded to verification items against the 'normal' grain). Verification items were not included in the analyses. The decision process to arrive at a factor structure was not as extensive due to fewer items and no emphasis on whether a combined or standalone dataset was used. The final factor structure for the anxiety scale is shown in **Table 4.9.** This structure was obtained using a pre-GC I dataset from Fall 2014.

Table 4.9. Component matrix: Four-factor solution for the anxiety scale from CSEAS (F14 GCI, N=110)

Factor & Item	Item	Four factors - Pre GC I			
		1	2	3	4
Factor 1	**Anxiety related to learning chemistry**				
1	Signing up for your next chemistry course	.823			
3	Learning chemistry in your current and future chemistry courses	.816			
4	Hearing the word "chemistry"	.758			
5	Learning new concepts in chemistry	.640			
Factor 2	**Anxiety related to low stakes assessments**				
11	Reading your chemistry textbook to help with homework		.843		
13	Watching and following your chemistry instructor work a problem on the board		.792		
12	Listening to lecture in your chemistry class		.767		
Factor 3	**Anxiety related to interactions**				
8	Talking to your chemistry professor			.845	
9	Asking or answering questions in your chemistry course			.790	
6	Walking into your chemistry lecture			.603	
Factor 4	**Anxiety related to high stakes assessments**				
14	Waiting to get a chemistry test returned				.930
7	Taking examinations in your current chemistry course				.511

Factor analyses – Confirmatory

As the overall concern of this project is to make meaningful measurements longitudinally, it was essential that the factor structure obtained by administering the CSEAS and the survey discussed in chapter 5 (COES) provided adequate to good model fit at time points that constituted the longitudinal period – the two-semester chemistry sequence of gateway courses, GC I and GC II. Thus, the factor structure shown in **Table 4.8** was imposed on self-efficacy datasets from each of these time points (pre-GC I, post-GC I and pre- GC II) to assess model fit and robustness of the factor structure. Model fit was evaluated with and without outliers resulting from the SAS code (included in **Appendix A**). In the case of CSEAS, as some of the items exhibited slightly high skew and kurtosis values (items 2 and 3), excluding outliers resulted in better model fit at all three time points.

The factor structure resulting from EFA did not provide adequate overall model fit, with RMSEA values close to 0.1. While modification indices were explored for recommendations to better model fit, other factor structures from EFA were also examined to see if items in their current subscales had been grouped differently (or more meaningfully) in other structures. Consequently, various models were attempted with some of the items placed in subscales other than their current ones. For instance, fit was assessed when two were grouped with factor 1 instead of factor 2. Reverting to EFA results of post-GC I, when the multitude of items in factor 1 were factor analyzed, they resolved into two factors; the same technique was attempted to assess model fit when factor 1 separated into the interpersonal subscale and remaining assessment items. After several iterations, the fit indices were still fair, with RMSEA values displaying good fit at two time points and CFI values well within range. The fit indices shown in **Table 4.10** have been obtained using datasets in which outliers (indicated by the output) were excluded. In addition, determining a meaningful factor for the CSEAS required the use of several trial datasets; consequently, CFA could only be conducted using combined data from two semesters. Only the most commonly reported fit indices are displayed in this table. Detailed description of the indices shown have been included in chapter 3.

Table 4.10. Goodness-of-fit indicators of models for CSEAS at three time points during AY15-16

Model tested on:	N	$\chi 2$	df	$\chi 2$ / df	SRMR	CFI	RMSEA	RMSEA CI
F15 + S16 pre -GC I	288	374.11	194	1.93	.054	.922	.060	0.0506 - 0.0688
F15 + S16 post-GC I	166	394.96	194	2.05	.068	.923	.079	0.0680 - 0.0904
F15 + S16 pre-GC II	201	350.61	194	1.81	.052	.939	.064	0.0528 - 0.0741

As shown in **Table 4.10**, the indices display good fit for pre-datasets from GCI and GCII; while the CFI and SRMR are well within range for the post GCI dataset, the RMSEA value represents a reasonable error of approximation with an upper limit on the confidence interval above the recommended threshold of .08. Given the loss of task specificity in subscales when examining pre vs. post factor structures in the CSEAS, it is possible that the degree to which the context/perception of a student affects their perceived confidence is much higher in a post scenario. The small sample size could also impact some indices relative to others. In addition, when timestamps outside the normal range were excluded, model fit worsened considerably, possibly suggesting that those who took longer to complete the survey may have taken the it more seriously. Thus, for datasets from all three time points, only responses with zero variance were excluded. Based on these observations, the model fit and periodic examination of the CSEAS factor structure are essential to ensure the stability of the model. The path diagram for the finalized CSEAS model is shown in **Figure 4.6**.

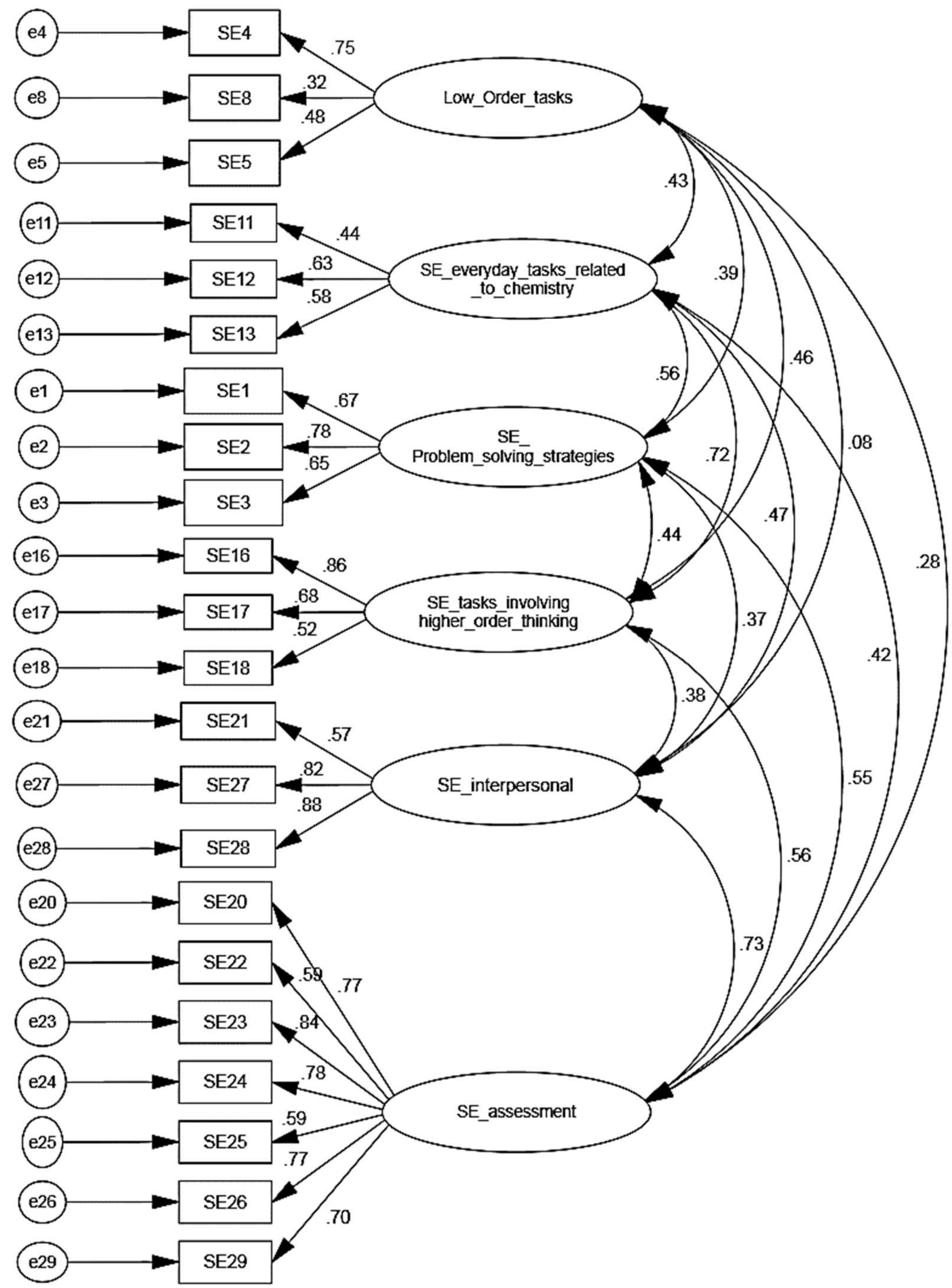


Figure 4.6. Standardized coefficients for the final, refined six-factor model of the chemistry self-efficacy and anxiety scale. All coefficients are significant at *p* <0.01.

Comparative statistics

Pre-post score changes on each subscale (factor) were examined using a GC I dataset. These results are shown in **Table 4.11**. A pre- and post-GC I sample was used as the test dataset for these analyses.

Table 4.11. CSEAS subscale scores showing pre to post changes for GC I (S14-S16, N = 367)

Factor	Avg. prettest scores	Avg. posttest scores	t	*p*	Effect size
1. SE related to assessment and evaluation	2.02	2.59	-10.211	<0.0001	0.669
2. SE related to interpersonal tasks	2.32	2.79	-7.866	<0.0001	0.506
3. SE related to applying strategies	2.23	2.04	4.199	<0.0001	-0.274
4. SE related to higher order tasks	2.70	2.31	7.828	<0.0001	-0.476
5. SE related to applying chemistry to everyday tasks	2.04	1.65	8.527	<0.0001	-0.545
6. SE related to low order tasks	2.14	1.88	5.970	<0.0001	-0.383

All six subscale scores showed significant changes across a semester, with self-efficacy related to applying strategies, performing low order tasks, higher order tasks and applying chemistry to everyday tasks showing lower average posttest scores, indicating higher self-efficacy (confidence). Subscales in which students displayed lower confidence at the end of the semester were those related to assessment and evaluation and interpersonal tasks. While confidence related to interpersonal tasks is fairly subjective, it is possible that most students taking their first college chemistry course might exhibit lower confidence with regards to interacting with their peers or instructors either in person or during their large lectures. Given the various components that contribute to students' assessment and evaluation, it would not be unusual for students to display lower confidence especially when the tasks in this subscale are being evaluated close to an impending assessment (final exam). Additionally, as stated during the interviews, if the mindset of hard work not translating to grades is pervasive, students could be drawing heavily on past experiences when responding to the items in this subscale. The indication of significant changes in each subscale suggests that there could be changes happening at key time points during the semester. While the task specific subscales displayed higher posttest confidence overall, examining these factors by subgroups might paint a different picture.

Differences between gender subgroups were examined for each subscale using a pre- and post-GC I dataset. Results for the pre-GC I dataset are shown in **Table 4.12**. For ease of interpretation, only subscales in which significant gender differences were observed are displayed.

Table 4.12. Results showing differences in CSEAS subscales based on gender (N_{males} = 153, N_{females} = 214), GCI pre, S14-S16.

Factor	Males	Females	t	*p*	Effect size, d_{Cohen}
2. SE related to interpersonal tasks	2.19	2.41	-2.537	.012	.270
5. SE related to applying chemistry to everyday tasks	1.94	2.11	-2.140	.033	.227

Subscale differences by gender (at the start of the course) reveal that females have lower self-efficacy than males with regards to interpersonal tasks and applying chemistry to everyday tasks despite being similar in ability levels (based on ACT test). As stated by female students in interviews, the interactions with their instructor depended to a large extent on the environment in which this interaction had to occur (lecture vs. in-person). Students mentioned that the they were fairly anxious interacting in a lecture setting as opposed to meeting their instructor during office hours. Given the minimal expectations about the course and perhaps their first experience with a college chemistry course, it is not unreasonable to expect students, especially females, to display lower confidence when it comes to interactions with their instructors or teaching assistants.

With regards to applying chemistry to everyday tasks, as the items comprising this task were fairly mathematical in nature (‘converting your speedometer reading’, ‘calculating the density of lemonade’), it is not unusual that females exhibited lower confidence in this subscale; the historically low self-efficacy of females in math-related tasks and their resulting avoidance of science-related careers is the seminal study (Hackett & Betz, 1989) that has shaped decades of work related to women in STEM fields. In addition, it is possible that some students started solving the problems or answering the questions within the subscale despite being instructed to indicate their perceived confidence. These in-the-moment problem solving attempts could have either worsened an existing low self-efficacy or triggered some anxiety if the students experienced difficulty with solving the problems. It is also possible that this subscale drew heavily upon students' past experiences, with females being more impacted by prior negative experiences.

When these same gender differences were tested for each subscale using a post-GC I dataset, there were no significant differences between males and females in each subscale. These results conflict with some other studies in literature, which have observed persistent gender differences across a semester. Regardless of the situation, examining gender differences in self-efficacy has produced mixed results, consequently requiring closer attention. Given the changes in social psychological climate and attributes of today's female scientists, the role of gender in self-efficacy and other affective measures requires further investigation. These results also justify the need for an instrument that can examine this construct on a finer level and identify the changes that are occurring during a semester.

Reliability

Factor correlations and factor alpha coefficients were calculated for the model confirmed by CFA. These results are shown in **Table 4.13**.

Table 4.13. Factor correlations and Cronbach alpha coefficients for the CSEAS (pre-GCI, S14-F16, N=453). All correlations are significant at the 0.01 level.

Factor	Mean	Std. dev.	1	2	3	4	5	6	Cronbach's alpha (α)
1. SE related to assessment and evaluation	2.08	.71	1.000						0.681
2. SE related to interpersonal tasks	2.31	.82	.611	1.000					0.736
3. SE related to applying strategies	2.27	.68	.451	.167	1.000				0.716
4. SE related to higher order tasks	2.72	.81	.296	.247	.263	1.000			0.712
5. SE related to applying chemistry to everyday tasks	2.09	.78	.298	.212	.396	.409	1.000		0.714
6. SE related to low order tasks	2.18	.68	.336	.194	.417	.418	.330	1.000	0.711
Total scale (n=22)									0.748

Reliability estimates ranged from .68 to .74 with a total scale Cronbach's alpha equal to .748. When the scale with 30 items was evaluated, Cronbach's alpha was equal to .905. Item-total statistics did not reveal low ITCs for the scale. The estimates overall and by subscale (except for factor 1) are above the recommended threshold of .70 (Nunally, 1978).

Validity

Construct validity was established using exploratory and confirmatory factor analysis. As the subscales established are fairly fluid and highly sensitive to the quality of standalone or combined data, especially given the administration using an online platform, these subscales were tested for fit using CFA with freshly collected data each semester. The presence of distinct subscales that show meaningful and significant changes across a semester is indicative of an instrument that is capturing a construct that is unlikely to be captured using performance indicators alone.

Predictive validity (a form of criterion-related validity) was evaluated by correlating mean subscale scores to placement test scores, final exam and course performance scores. These analyses were conducted using a pre and post GC-I dataset; the pre-GC I correlations are displayed in **Table 4.14.**

Table 4.14. Correlations between pre-CSEAS subscale scores and performance indicators in GC I (N = 302, S14-S16). **. Correlations significant at 0.01 level; *. Correlations significance at 0.05 level.

Factor	ACT Comp.	ACT Math	ACT-SciRe	Final Exam %	Course %
1. SE related to assessment and evaluation	-.087	-0.142*	-.086	-.250**	-.296**
2. SE related to interpersonal tasks	-.046	-.046	-.068	-.119*	-.197**
3. SE related to applying strategies	-.124*	-.151*	-.079	-.246**	-.264**
4. SE related to higher order tasks	-.003	0.03	-.013	-.102	-.114*
5. SE related to applying chemistry to everyday tasks	-.122*	-.206*	-.078	-.276**	-.304**
6. SE related to low order tasks	.069	.044	.064	-.120*	-.159**

Research on the relationship between self-efficacy and achievement has demonstrated significant and positive correlations (*r* values in a range of 0.38 to 0.42) between self-efficacy for cognitive abilities or skills, which are assessed prior to instruction (Schunk & Hanson, 1985); positive correlations in the range of *r*=.46 to .90 have also resulted between self-efficacy and skill assessed after instruction (Schunk, 1989; Schunk, 1995). In addition, self-efficacy and performance, evaluated after instruction, have consistently displayed significant and positive correlations in the range of *r*=.27 to .84 (Schunk, 1989).

The results in **Table 4.14** offer support for some of these relationships; the correlations are negative because in the CSEAS, a high average self-efficacy score implies low self-efficacy (confidence). Self-efficacy related to applying strategies and chemistry to everyday tasks (mathematical problem solving tasks) show significant 'positive' correlations to ACT composite and math scores. As these were evaluated prior to instruction or interaction, confidence related to interpersonal tasks showed no significant correlations with any measures of cognitive abilities. The tasks most relevant and similar to what students' might have encountered to test their abilities are tasks related to applying general strategies and problem solving tasks (which comprise the subscale related to applying chemistry to everyday tasks). This could perhaps be the reason behind these subscales 'positively' correlating with ACT composite and math scores.

Perceived confidence in all pre-subscales significantly correlates with students' performance in the course. The results of this correlational analysis using post-CSEAS subscale scores is shown in **Table 4.15**.

Table 4.15. Correlations between post- CSEAS subscale scores and performance indicators in GC I (N = 302). **. Correlations significant at 0.01 level; *. Correlations significance at 0.05 level.

Factor	ACT Comp.	ACT Math	ACT-SciRe	Final Exam %	Course %
1. SE related to assessment and evaluation	.011	-.071	-.031	-.163**	-.260**
2. SE related to interpersonal tasks	.034	-.007	.019	-.035	-.121**
3. SE related to applying strategies	-.041	-.095	-.108	-.111	-.177**
5. SE related to applying chemistry to everyday tasks	-.106	-.101	-.124*	-.178**	-.193**
6. SE related to low order tasks	-.050	-.035	-.041	-.088	-.113*

The correlations between CSEAS subscales and course performance, using post subscale scores, are significant, but weaker than with pre-subscale scores. Self-efficacy related to higher order tasks did not significantly correlate with any performance indicators. Given that these results are post instruction, perceived confidence levels are fairly conservative, perhaps realistic, and align with the lower end (closer to r=.27) of the range of correlations mentioned by Schunk (1985). In addition, while almost all subscales correlate significantly with course performance, self-efficacy related to assessment and evaluation and applying chemistry (mathematical problem solving) are significantly correlated to the performance on the final exam. Given that these two subscales are assessing perceived self-efficacy beliefs in tasks that closely correspond to the criterial task to which they are compared, a positive relationship would be expected between these subscales and the final exam. On the other hand, course performance, dependent on performance on several tasks, is correlated positively to almost all subscales, which involve tasks that contribute in some way towards successful performance in the course. While these correlations imply only weak to moderate relationships, these are indications that the CSEAS is capturing affective dimensions and is not just another measure of academic ability.

When differences were examined by high vs. low performing student groups based on final exam performance, significant differences were observed between both groups on all subscales with high performing students consistently displaying lower average subscale scores, implying higher self-efficacy (confidence) than the low performing student group. These results are shown in **Table 4.16**.

Table 4.16. Results showing differences in subscales based on high vs. low performing groups on the final exam - CSEAS ($N_{\text{low peformers}}$ = 172, $N_{\text{high performers}}$ = 196) – GC I pre, S14-S16

Factor	Low performers	High performers	t	*p*	Effect size
1. SE related to assessment and evaluation	2.20	1.86	4.784	<0.0001	-0.500
2. SE related to interpersonal tasks	2.42	2.23	2.162	0.031	-0.226
3. SE related to applying strategies	2.43	2.05	5.599	<0.0001	-0.585
4. SE related to higher order tasks	2.81	2.61	2.374	0.018	-0.248
5. SE related to applying chemistry to everyday tasks	2.26	1.84	5.378	<0.0001	-0.565
6. SE related to low order tasks	2.27	2.03	3.468	0.001	-0.362

When differences were further examined by gender and performance on the final exam, no significant differences were observed between high performing males vs. females on any of the subscales. However, when low performing students were examined, significant differences were found between low performing males and females on two subscales – self-efficacy related to interpersonal tasks and self-efficacy related to low order tasks; low performing females exhibited lower efficacy than males with regards to both subscales. These results are detailed in **Table 4.17**.

Table 4.17. Results showing gender differences in subscales based on low performance on the final exam – CSEAS (N_{males} = 52, $N_{females}$ = 119); GC I pre, S14-S16.

Factor	Males	Females	t	*p*	Effect size
2. SE related to interpersonal tasks	2.22	2.52	-2.146	0.032	0.359
6. SE related to low order tasks	2.09	2.34	-2.088	0.038	0.347

When this process was repeated using performance in the course, significant differences were observed between high vs. low performing groups on all subscales except self-efficacy related to higher order tasks. These results are displayed in **Table 4.18**. Only subscales in which significant differences were observed are shown.

Table 4.18. Results showing differences in subscales based on high vs. low performing groups in the Course - CSEAS ($N_{low\ peformers}$ = 168, $N_{high\ performers}$ = 200); GC I pre, S14-S16.

Factor	Low performers	High performers	t	*p*	Effect size
1. SE related to assessment and evaluation	2.21	1.86	4.894	<0.0001	-0.512
2. SE related to interpersonal tasks	2.46	2.21	2.877	0.004	-0.302
3. SE related to applying strategies	2.41	2.08	4.792	<0.0001	-0.501
5. SE related to applying chemistry to everyday tasks	2.25	1.84	4.961	<0.0001	-0.527
6. SE related to low order tasks	2.23	2.06	2.417	0.016	-0.225

High performing students displayed higher confidence than low performing students on the subscales that showed significant differences. With regards to the non-significant differences for self-efficacy related to high order tasks, it is possible that confidence in performing higher order tasks or demonstrating divergent thinking were perceived as crucial to a high stakes assessment such as the final exam; however, as performance in the course was dependent on a multitude of tasks throughout the semester with not all tasks requiring a higher level of thinking (for example, moodle homework or tasks that involved study groups for lab reports or discussion activities), this subscale did not display significant differences between high vs. low performing student groups based on course performance.

When differences were further examined by gender and course performance, no significant differences were observed between high performing males vs. females on any of the subscales. However, when low performing students were examined, significant differences were found between low performing males and females on two subscales – self-efficacy related to interpersonal tasks and self-efficacy related to low order tasks; low performing females exhibited lower efficacy than males with regards to both subscales. These results are detailed in **Table 4.19**.

Table 4.19. Results showing gender differences in subscales based on low performance in the Course - CSEAS (N_{males} = 61, $N_{females}$ = 106); GC I pre, S14-S16

Factor	Males	Females	t	*p*	Effect size
2. SE related to interpersonal tasks	2.29	2.56	-2.054	0.046	0.330
6. SE related to low order tasks	2.06	2.33	-2.370	0.014	0.381

Evidence of convergent validity was provided by examining relationships between the self-efficacy and anxiety subscales of the CSEAS. These correlations are displayed in **Table 4.20**. The correlations are positive because a higher average subscale score on the self-efficacy scale implies lower self-efficacy (confidence) while a higher average subscale score on the anxiety scale implies higher anxiety. It was expected that anxiety related to high stakes assessment would correlate strongly with self-efficacy related to assessment and evaluation. In addition, anxiety related to interactions would be expected to correlate strongly with self-efficacy related to interpersonal skills. Moreover, it was also expected that anxiety related to low stakes assessment would correlate strongly with self-efficacy related to low order tasks.

Table 4.20. Correlations between self-efficacy and anxiety subscales in the CSEAS – GC I (F14-F15, N = 452). **Correlation is significant at the 0.01 level; *Correlation is significant at the 0.05 level

Factor	Anxiety related to learning chemistry	Anxiety related to low stakes assessments	Anxiety related to interactions	Anxiety related to high stakes assessments
1. SE related to assessment and evaluation	.561**	.461**	.419**	.358**
2. SE related to interpersonal tasks	.289**	.320**	.523**	.275**
3. SE related to applying strategies	.320**	.248**	.164**	.171**
4. SE related to higher order tasks	.169**	.141**	.061	.070
5. SE related to applying chemistry to everyday tasks	.180**	.162**	.144**	.118*
6. SE related to low order tasks	.221**	.190**	.081	.039

As indicated in the **Table 4.20**, self-efficacy and anxiety subscales show moderate to strong 'negative' correlations, supporting their inverse relationship. As anxiety related to learning chemistry increases, self-efficacy related to assessment decreases. The weaker correlation between anxiety related to high stakes assessments (taking examinations and waiting to get a test returned) and self-efficacy related to assessment and evaluation could perhaps suggest the highly situational and contextual nature of anxiety and self-efficacy. While students may perceive 'taking examinations' as threats and not feel confident about being in a test-taking environment (anxious test takers), learning chemistry not only involves a greater contribution from the student in terms of self-regulatory behavior and implementing effective study habits, but it also incorporates connecting knowledge from various components and materials available in the course. A high anxiety level with regards to learning chemistry could, in the absence of help, result in poor study habits and decrease self-efficacy (related to assessment and evaluation) to a much greater extent than a high anxiety level with regards to high stakes assessments. The strong relationship between anxiety related to interactions and self-efficacy related to interpersonal tasks offers support for convergent validity; one would expect these subscales to correlate strongly as they are measuring common dimensions between two distinct constructs. The non-significant relationship between anxiety related to interactions and self-efficacy related to higher order tasks is not unusual given that there is minimal commonality between two fairly different subscales.

Further validation was provided when average subscale scores were correlated to other items in the CSEAS. When subscale scores were correlated with the importance of chemistry in students' academic preparation (1=very important to 5=very unimportant), there were significant, positive correlations observed for self-efficacy related to low order tasks (r=.158), self-efficacy related to applying chemistry to everyday tasks (r=.151), self-efficacy related to applying strategies (r=.133), self-efficacy related to higher order tasks (r=.122) and self-efficacy related to assessment and evaluation (r=.163). These findings corroborate results from Hackett and Betz's (1983) administration of the MSES, in which they indicated that students with stronger mathematics self-efficacy expectations had a greater tendency to view math as useful.

As a final check of validity, results from students' interviews were examined for student generated item groupings. An example of item groupings and group names generated by a female interviewee is shown in **Table 4.21**. This student was a kinesiology major (declared) who had previously taken three physics courses and her experiences in these courses impacted the way she responded and perceived the CSEAS items. These students were given all 30 items to sort into groups; no item was excluded.

Table 4.21. Example of student generated CSEAS item groupings and subscale names – GC I

	Item groupings
application of chem. to everyday tasks	Describing fundamental structure of an atom (7) Learning material in chemistry courses where considerable math is involved (22) Using chem. to propose solution that keeps cooking water from boiling over (17) Classifying Al foil, salt & salad dressing as elements, mixtures, compounds (15) Temperature in your home from degree Fahrenheit to K (9) Identifying type of change (physical vs. chemical) when milk goes sour (13) Writing summary of the main pts of a TV docu. that deals with some aspect of chem (18) Explaining why addition of salt melts ice (16) Percent composition of iron in 10g rust (Fe_2O_3) from garage door (14) Writing the chemical formula of Calcium carbonate, TUMS ingredient (10)
On its own	**Signing up for more chemistry courses in the future (regardless….) - 25**
Mastery of skill study guide/exam	Density of lemonade (50g lemons to 500mL water) - 12 Converting speedometer reading from mph to yards/sec (1 mile = 1760 yds) (11) Balancing chemical equations (6) Describing the trends in the periodic table (atomic size, electroneg.) - 5 Choosing an appropriate equation to solve a chemistry problem (2) Determining the appropriate units for a numerical result (3) Elements that are gases at rtp (from periodic table) - 8 Reading and writing a chemical formula (4)
Background (sp. To course) person + effort	Learning chemistry in this course (if all exams were take home exams) - 19 Understanding your chemistry professor - 27 Talking to your chemistry professor - 28 Asking questions during lecture - 21 Doing well on chemistry course exams, given you exert effort (20) Understanding what a written chemistry problem is asking you to do (1)
Actual exam subcategory of 3rd group	Taking an exam or quiz in your chemistry course - 23 Doing homework for this course - 30 Receiving grade you desire in this course (29) Preparing for chemistry exams - 26 Taking a chemistry exam or quiz where considerable math is involved - 24

Most female students went in-depth and created highly specific groups such as:

- Post exam (items 20, 29, 25)
- Related to exam (items 1,2,23,24)
- Before class and outside (items 26,30)
- Tasks that would benefit their grades (items 21,28,27,19)
- Everyday life (items 16, 17, 15, 9, 18, 14, 11, 12)
- Interaction (21,27,28)

The indication that some students perceived interaction or interpersonal tasks as among those that would benefit their grades helps to explain the factor structures in which assessment and interpersonal tasks would combine to form one factor. In addition, despite the presence of a factor that was categorized in the study as 'higher order tasks', students seldom perceived tasks in terms of the level of thinking that would be required to perform the task. While female interviewees parsed out tasks that involved math vs. other basic (general) material, male students did not perceive any degree of task specificity in the survey items. An engineering student in GC I, who was exploring other majors, categorized the items as follows:

- Interaction (items 21, 28, 27)
- Learning throughout the course (items 3,4,5,2,1,6,7,8,9,10,11,12,13,14,15,16,17,18,19,22,30)
- Exams/assessment (items 23,26,24)
- End results/evaluation (items 20,29,25)

While the comparisons were not meant to be identical, it was observed that most students were essentially thinking about items in a similar context. The lack of task specificity in groups for a few interviewees might explain some of the correlation results in this study. If students were perceiving all non-assessment related items as one task group, it would be expected that all the subscales related to chemistry tasks would be correlated in a similar manner to key performance indicators. Given that students were able to create groups to begin with was an indication that items were factoring either because students were using the same or similar context association or that the association was much less important than the perception of problem solving separate from context.

Cluster analyses

The cluster structures resulting from data collected in preparatory chemistry and general chemistry for engineers are shown in **Tables 4.22** and **4.23**. As detailed earlier, the cluster structures are displayed here only in an effort to examine utility of the CSEAS instrument. No items were excluded from the analyses.

<u>Preparatory chemistry</u>

The dataset used for obtaining the cluster structure consisted of 668 students, out of whom 41.9% were males and 58.1% were females. 52.5% were freshmen, 42.4% were sophomores, 12.4% were juniors and 7.6% were seniors. The average ACT composite, math and sci-re scores were 22.4, 22.3 and 22.6 respectively. The most meaningful cluster structure was obtained using the 30-item survey when items 19 and 20 were excluded. This structure is shown in **Table 4.22**.

<u>General chemistry for engineers</u>

The dataset used for obtaining the cluster structure consisted of 238 students, out of whom 91.6% were males and 8.4% were females. 37.8% were freshmen, 38.7% were sophomores, 13.0% were juniors and 10.5% were seniors. The mean ACT composite, math and science-reasoning scores were 24.2, 25.3 and 24.6 respectively. The most meaningful cluster structure was obtained using the 30-item survey; items 9 and 30 were excluded. This structure is shown in **Table 4.23**.

Table 4.22. Three-cluster solution for preparatory chemistry (N=668)

Cluster and item	Item
Cluster 1	**Self efficacy related to applying chemistry to everyday tasks**
16	Explaining why addition of salt melts ice
17	Using chemistry to propose a solution that keeps cooking water from boiling over
14	Calculating the % composition of iron in rust (Fe2O3) from garage door
13	Identifying the type of change (physical vs. chemical) when milk gets sour
15	Classifying aluminum foil, salt and salad dressing as compounds, mixture or elements
12	Calculating the density of lemonade (made by adding 50g of lemons to 500mL of water)
10	Writing the chemical formula of Calcium carbonate, TUMS ingredient
18	Writing a summary of the main points of a television documentary that deals with some aspect of chemistry
11	Converting your speedometer reading from mph to yards/second (1 mile = 1760 yards)
9	Converting the temperature in your home from F to Kelvin
Cluster 2	**Self efficacy related to assesment and evaluation**
29	Receiving the grade you desire in this course
26	Preparing for chemistry exams
23	Taking an exam or quiz in your chemistry course
28	Talking to your chemistry professor
27	Understanding your chemistry professor
21	Asking questions during lecture
24	Taking a chemistry exam or quiz where considerable math is involved
25	Signing up for more chemistry courses in the future (regardless of the outcome of this course or the requirements for your major)
22	Learning material in chemistry courses where considerable math is involved
30	Doing homework for this course
Cluster 3	**Self efficacy related to general chemistry tasks**
2	Choosing an appropriate equation to solve a chemistry problem
1	Understanding what a written chemistry problem is asking you to do
8	Identifying elements that are gases at room temperature (from the periodic table)
7	Describing the fundamental structure of an atom
5	Describing the trends in the periodic table (atomic size, electroneg.)
4	Reading and writing a chemical formula
6	Balancing chemical equations
3	Determining the appropriate units for a numerical result

Table 4.23. Three-cluster solution for general chemistry for engineers (N = 238)

Cluster and item	Item
Cluster 1	**Self efficacy related to assesment, evaluation and interpersonal tasks**
1	Understanding what a written chemistry problem is asking you to do
2	Choosing an appropriate equation to solve a chemistry problem
3	Determining the appropriate units for a numerical result
21	Asking questions during lecture
22	Learning material in chemistry courses where considerable math is involved
23	Taking an exam or quiz in your chemistry course
24	Taking a chemistry exam or quiz where considerable math is involved
25	Signing up for more chemistry courses in the future (regardless of the outcome of this course or the requirements for your major)
26	Preparing for chemistry exams
27	Understanding your chemistry professor
28	Talking to your chemistry professor
29	Receiving the grade you desire in this course
Cluster 2	**Self efficacy related to general chemistry tasks**
4	Reading and writing a chemical formula
5	Describing the trends in the periodic table (atomic size, electroneg.)
6	Balancing chemical equations
7	Describing the fundamental structure of an atom
8	Identifying elements that are gases at room temperature (from the periodic table)
10	Writing the chemical formula of Calcium carbonate, TUMS ingredient
11	Converting your speedometer reading from mph to yards/second (1 mile = 1760 yards)
12	Calculating the density of lemonade (made by adding 50g of lemons to 500mL of water)
13	Identifying the type of change (physical vs. chemical) when milk gets sour
14	Calculating the % composition of iron in rust (Fe2O3) from garage door
Cluster 3	**Self efficacy related to low stakes tasks**
15	Classifying aluminum foil, salt and salad dressing as compounds, mixture or elements
16	Explaining why addition of salt melts ice
17	Using chemistry to propose a solution that keeps cooking water from boiling over
18	Writing a summary of the main points of a television documentary that deals with some aspect of chemistry
19	Doing homework in this course
20	Doing well on chemistry course exams, given you exert enough effort

The cluster structure of data from preparatory chemistry was similar to the item groups generated by GC-I students during interviews. The scope of task specificity was grouping items as general chemistry tasks or those which involved applying chemistry to everyday tasks. Items 19 and 20 were excluded from analyses as they either formed a group together, appeared as standalone items or did not fall into a meaningful cluster. Given the highly fundamental level of knowledge of students in this course, it is possible that students were unfamiliar with the types of problems or questions consequently responding to items from a problem-solving perception as opposed to using contextual associations.

A similar structure is obtained by analyzing data from general chemistry for engineers. Items 9 and 30 were excluded as they always appeared as standalone items. While most of the general chemistry task items grouped together, cluster 3 was home to some items that did not appear to substantively belong in that cluster. In this structure, the omnibus items and interpersonal tasks were perhaps interpreted as tasks that had to be performed to ensure success in the course; consequently, they clustered with assessment related items.

Despite being a feeder and terminal course respectively, the structures resulting from these courses appear to be meaningful. The lack of task specificity in content areas from both structures suggests that students may be placing less emphasis on contextual associations and more on the perception of problem solving. Despite this, the formation of distinct groups comprising of items with a moderate to high degree of relatedness suggests that the CSEAS could be a viable survey to measure self-efficacy in courses besides general chemistry.

Limitations

Obtaining a factor structure for this survey was a not so easy quest especially as this was previously validated. Being a follow-up instrument to an extensive concept inventory, the ongoing concern was always one of student fatigue resulting in loss of valid responses. While factor structures from standalone and sequential administration of surveys have been examined, with no major differences, making affective measurements using an online platform was expected to be prone to problems regarding missing data, ceiling/floor effects and students clicking responses with little to no thought behind their choices. As all the items in the CSEAS were positively worded, the chance of acquiescence bias was high. Thus, zero variance items were always excluded. However, despite data checks and monitoring of timestamps in the collected data, online self-report surveys are unlikely to be completely free of bias.

While the items in this survey have attempted to include a reasonable degree of specificity in terms of content and course, and are purported to measure self-efficacy related to chemistry vs. general self-efficacy, it is possible that a certain degree of the latter might be unavoidably incorporated in student responses.

The analysis in this study, especially obtaining and confirming a factor structure, was limited by the size of some standalone datasets such as pre-GC II. Thus, the analysis had to be conducted using a combined dataset to begin with, thereby not allowing for comparisons to be made between results obtained using standalone vs. combined datasets. In addition, while some of the items used in this survey (items 1,2 and 3) have been criticized as omnibus measures that assess general self-efficacy and have low predictive power (Pajares, 1996), they were utilized in this survey as problem solving strategies applicable to several domains, including chemistry. Although the factor structure resulting from data collected using this survey is being checked periodically and with larger datasets, these results appear to be generalizable to the same institution / student type.

Conclusions and Implications

This study presented a detailed description of the phases involved in developing/adapting and validating an instrument to measure self-efficacy. In an effort to aid in psychometric testing, an anxiety survey was also developed and integrated with the self-efficacy component to result in the chemistry self-efficacy and anxiety survey (CSEAS). The pilot administration of this survey was done on paper, with subsequent administrations online. EFA of the 30-item instrument resulted in exclusion of eight items and a six-factor factor whose fit was tested using CFA. While this model displayed good fit at pre-GC I and pre-GC II time points, the post-GC I time point resulted in a poor fit. This could be attributed to the lack of task specificity in the post factor structure. The differences between self-efficacy for learning and self-efficacy for performance might explain the lack of task specificity in post-factor structures. Given that students become familiar with tasks at the end of the semester, their post self-efficacy beliefs are based on their past efficacy beliefs in similar tasks during the semester. As their post-beliefs are closely tied to performance, it is possible that students view their confidence in performing all tasks, including assessment related ones, as being important towards their intended performance in the course. On the contrary, pre self-efficacy beliefs are more inferential and less inclined to be drawn from past experiences; thus, students might be responding to pre-surveys based on the idea that their perceived confidence in specific tasks will help them accomplish similar tasks they might encounter during the semester. Cronbach's alpha for the 22-item scale was .748 with reliability estimates ranging from .68 to .74. All six subscales showed significant changes across a semester, indicative of an instrument that was capturing dimensions of a variable that could not be measured using traditional performance indicators. Validity for the CSEAS was supported by correlational analyses, comparative statistics and student interviews.

All pre-CSEAS subscales were significantly correlated with course performance while all subscales except self-efficacy related to higher order tasks showed significant correlations with performance on the final exam. These correlations were low to moderate, suggesting that the CSEAS was not simply another measure of academic performance. The post-subscale coefficients, although weaker, showed significant correlations of all subscales with performance in the course. The presence of a significant relationship between self-efficacy and performance offers support for the domain specificity or correspondence of the CSEAS. Gender-based differences were observed for self-efficacy related to interpersonal tasks and self-efficacy related to applying chemistry to everyday tasks. Women exhibited lower confidence than men in both subscales. These differences were no longer significant in any subscale at the end of the semester.

When parsed out by performance, high performing students on the final exam consistently displayed higher confidence on all subscales than low performers. When this was examined by performance in the course, high performing students were more confident than low performing students on all subscales except self-efficacy related to higher order tasks. When high performing students were parsed out by gender, there were no significant differences between high performing males or females on any subscale. However, when examining low performers, significant differences were found between males vs. females, with females displaying lower confidence than males on two subscales – self-efficacy related to interpersonal tasks and self-efficacy related to low order tasks.

Furthermore, self-efficacy subscales were correlated with anxiety subscales to reveal strong 'negative' correlations between almost all anxiety and self-efficacy subscales. The strongest correlations were among self-efficacy related to assessment and evaluation and all the anxiety subscales. Self-efficacy subscales also showed positive correlations with the importance that students placed on chemistry in their academic preparation. The ability of the CSEAS to measure fairly distinct clusters in preparatory chemistry and general chemistry for engineers provide support for the viability of this instrument to make meaningful self-efficacy measurements in courses related to general chemistry.

The CSEAS was partly developed and mostly adapted from the vast number of valid and reliable self-efficacy surveys in the literature. While validity in the CSEAS was purported based on the general acceptance of its constituent parts by numerous efficacy researchers, psychometric testing of this survey was conducted to establish validity and reliability for its use in the two-semester general chemistry sequence of courses that constitute time points within the longitudinal model; the utility of this survey was also tested on students in preparatory chemistry and general chemistry for engineers. This instrument complements the existing body of self-efficacy research; while several open ended and contextual items of this survey were unused for testing, they present substantially beneficial information that could be used to understand the self-efficacy and behaviors of students enrolled in chemistry courses. The task specific factors obtained and confirmed in this study allow for the development of a subset instrument that will be utilized to track potential changes in self-efficacy during the semester.

CHAPTER 5: DEVELOPMENT AND VALIDATION OF THE CHEMISTRY OUTCOME EXPECTATIONS SURVEY

This chapter describes the development and psychometric evaluation of data produced by the Chemistry Outcome Expectations Survey (COES).

Background and Rationale

Despite the integration of outcome expectations as a distinct construct in SCCT and the combined role of self-efficacy and outcome expectations in predicting interests, this construct has not received the same attention as its more prevalent companion construct, self-efficacy. From an assessment perspective, the domain specific nature of OE has resulted in varied operationalizations of the construct. Brooks and Betz (1990) developed an instrument to measure occupational values, but each item in the instrument was examined individually as opposed to using a summation of the items. Riggs et al., (1994) developed the Personal Outcome Expectancy Scale to measure OE for individuals in the workplace. Hackett et al., (1992) developed the Outcome Expectations Scale to assess how successfully people complete a bachelor's degree in engineering. The Educational Outcome Expectancy Scale (EOE), developed by Springer et al., (2001) is a six-item Likert-type scale to capture the consequences an individual expects from completing a bachelor's degree. However, six items were inadequate to measuring the full range of OE, thereby limiting the utility of this scale. The Vocational Outcome Expectations (VOE) scale has been used to measure expectations about future career outcomes. Using SCCT as a framework, this scale was developed to capture youths' sense of being able to obtain a fruitful vocational outcome (McWhirter, Crothers, & Rasheed, 2000).

In an effort to comprehensively measure outcome expectations, Smith & Fouad (1999) developed an instrument comprising of 153 items rated on a six-point Likert scale. The items measure self-efficacy, outcome expectations, interests and goals in four subject matter areas: math/science, social studies, English and art. In the physical sciences, the Maryland Physics Expectations Test (MPEX), a five-point Likert-scale survey was developed to measure students' attitudes about physics and their cognitive expectations of physics courses (Redish, Steinberg & Saul, 1998). The MPEX consists of 34 items grouped into six clusters that represent different aspects of expectations about learning: independence, coherence, concepts, reality link, math link, and effort (Redish, Steinberg & Saul, 1998). In addition, a more widened gap between student and faculty expectations was observed over the course of a semester. In an effort to investigate the existence of a similar phenomenon in chemistry, Grove & Bretz (2007) developed the Chemistry Expectations Survey (CHEMX), consisting of 47 statements grouped into seven clusters: Effort, concepts, math link, reality link, outcome, laboratory, and visualization. Contrary to merely changing survey domain from physics to chemistry, development of this instrument integrated various concepts and dimensions specific to chemistry such as visualizations and a focus on the particulate, symbolic and macroscopic aspects. Out of 47 statements, 22 were original items written specifically for CHEMX (Grove & Bretz, 2007). While these surveys were used to gauge attitudes and cognitive expectations in physics and chemistry courses, they did not necessarily probe the construct of outcome expectations as the statements were not operationalized in the context of predictive "if-then" statements.

Given the limitations of these surveys and the resulting inadequacies in capturing outcome expectations meaningfully, the following objectives guided this study:

a) To develop an instrument to assess chemistry outcome expectations.

b) To establish validity and reliability of the data resulting from this instrument.

Methodology

This section describes the phases involved in the development of the COES. The selection of items, construction of the instrument, testing and participants will also be detailed. In addition, the analyses conducted to psychometrically evaluate and validate the data will also be examined.

Development of the COES items

Operationalizing OE for development of the COES required the fulfillment of a few criteria:

a) Conceptually distinguishing it from self-efficacy – The focus of outcome expectations was not on the behavioral performance itself, but rather on the likely consequences of the behavior. Thus, the 'predictive' nature of this construct had to be reflected in the items, which were consequently expressed as 'if-then' statements (Fouad & Guillen, 2006). In distinguishing these two constructs, the strength of the relationship between them had to be taken into consideration when formulating statements in the COES:

(i) Complete relationship of OE to SE: As outcomes people expect are largely dependent on their self-efficacy beliefs, it is quite possible that the outcomes are completely dependent on SE beliefs. Thus, some of the tasks / behaviors (condition statements) in the COES were complements of the tasks from the CSEAS, for example "confidence in <u>understanding one's chemistry professor"</u> would be complemented by "If I can <u>follow my instructor in lecture</u>, then…" in the COES. This relationship has also been supported by the self-efficacy-antecedent model, which tested the temporal path of self-efficacy being a precursor of outcome expectations, interests and goals. Results from this investigation offered support for SCCT's hypothesis that the predominant temporal path was from self-efficacy to other variables rather than vice versa (Lent, 2008).

(ii) Partial relationship of OE to SE: When outcomes that result from specific performances are not themselves controlled by such performances, efficacy beliefs account for a smaller part of the variance in outcome expectations (Bandura, 1986).

(iii) Completely excluding considerations of outcome from self-efficacy judgments.

b) Domain specificity – Being a domain-specific construct, the survey statements were tailored to measure student outcome expectations in chemistry vs. general expectations about their program or expectations in other domains.

c) Integrating different forms of outcome expectations – According to Bandura (1997), outcome expectations can take three major forms – physical outcomes of the behavior, social reactions to the behavior and self-evaluative reactions to personal behavior. Physical outcomes include the pleasant physical sensations that follow behaviors, social reactions – positive and negative – are the second form of behavioral outcomes, where positive social reactions include approval and recognition while negative reactions include disapproval and social rejection. Self-evaluative reactions – positive and negative – are the third form of outcomes that accompany behaviors. Positive and negative self-evaluations include self-satisfaction and self-criticism respectively (Bandura, 1997; Fouad & Guillen, 2006). While the predictive utility of these specific types of expectations has not been examined, the statements in the COES have attempted to capture these forms of outcome expectations, in addition to investigating the effects of proximal (course) and distant (career) outcomes. Items related to career outcome expectations were included as a way to understand career indecision / career choices as one progressed through their major. In addition, these outcome expectations were viewed as perceived environmental contingencies – outcome expectations beyond an individual's control and independent of one's self-perception of competence (Bandura, 1986).

The item pool for the development of the COES was obtained from an extensive literature review of existing surveys attempting to measure outcome expectations in STEM and non-STEM domains. As the items in these surveys were declarative statements, they had to be adapted in ways that would make them usable as conditions or outcomes for the COES. In addition, if these statements were used as conditions, consequences or outcomes had to be added and vice versa. The items specific to chemistry were adapted primarily from ChemX and to a small degree from its predecessor MPEX. Once items were gathered from all surveys, they were grouped based on what each item was assessing and the utility of each item towards measuring some aspect of outcome expectations. The first step, investigating content validity, involved refining the initial pool of items; this was done by the principal investigator (PI) and graduate student working on this project. The decision to reduce the number of items to a manageable, yet meaningful instrument was made by (a) using gaps in the literature to dictate the utility of the COES and (b) using information based on the PI's teaching experiences, interactions with students and their posited interpretation of the items. Three examples of what this process entailed are described in **Table 5.1**, where an original item from the ChemX was refined and tailored to the COES. The change made to the original item is indicated (in bold).

Table 5.1. Examples of revisions made to items from ChemX for COES item development

Example item	Original item (ChemX)	Revision	Revised item(s) (COES)
1	It is possible to pass this course (get a "C" or better) without understanding the chemistry well	**Made more specific**	If I do not understand the concepts in this course, I can pass (with at least a C)
2	The main skill I get out of this course is to learn how to reason logically about the physical world.	**Added consequence**	If I can learn how to think logically about the physical world, I will do well in this course
3	A good understanding of chemistry is necessary for me to achieve my career goals. A good grade in this course is not enough	**Split into two statements**	i) If I have a good understanding of chemistry, I will have a better chance of achieving my career goals AND ii) If I obtain a good grade in this course, I will have a better chance of achieving my career goals

Statements from other surveys which measured career outcome expectations that pertained to obtaining a degree or a job were also modified into "if-then" statements.

Subsequently, in an effort to reduce acquiescence bias, some of the OE statements were negatively worded. As these were "if-then" statements, a decision had to be made about which part of the statement – task or outcome – would be negatively worded. In this case, the purported relationship between self-efficacy and outcome expectations had to be considered to arrive at a decision. As all the tasks described in the CSEAS items were positively worded, it was decided that tasks in some of the COES items would be negatively worded.

Structure of the COES

The COES started with four Likert-type statements from the College Student Experiences Questionnaire (CSEQ); the goal of the CSEQ was to assess new student decisions and expectations about how and with whom they will interact in college and how much effort would be invested in using institutional resources (Pace & Kuh, 1998). Thus, some of the items regarding 'time spent' were modified for use in the context of career development and included in the COES in an attempt to examine how dimensions of the COES may relate to student intentions or career exploratory plans. For example, statements in the CSEQ pertaining to time spent on learning material in the course could be related to statements in the COES that were relevant to outcomes expected from learning material in the course. The prompt for these statements was "How often do you expect to do the following?" with students responding with the amount of time (1 = never to 4 = very often and 5 = not sure). As this survey (CSEQ) operation closed in Spring 2014, the full survey is no longer available online. This survey is included in **Appendix I.**

This was followed by the Likert-type COES "if-then" items which were chosen for the final version of the survey based on (a) results from pilot testing the survey (b) discussions between principal investigator and graduate student and (c) semi-structured student interviews. As this survey was administered in a pre/post manner to measure outcome expectations, some statements in the post version of the survey were modified to allow students to respond to prospective outcomes (future chemistry course) rather than have them retroactively make causal associations between the task and outcome, especially when the outcome was tied to performance in the course. An example of this change from pre- to post-COES is shown below:

1A) Pre- COES: If I work hard enough, I will be more likely to pass this course.

1B) Post- COES: If I work hard enough, I will be more likely to pass a future chemistry course.

The pilot version of the pre-COES had 28 items, many of which were revised and/or removed for the post version based on student interviews and analyses.

Student interviews – instrument development and implementation

The solicitation and interview processes were similar to those described in chapter 4 for the CSEAS. 13 students signed-up to participate in semi-structured think aloud interviews, during which the survey was presented and students responded to each statement while verbalizing their thought processes. Almost all students were biology majors on a pre-med track with an equal number of male and female participants. Compensation was a $10 gift card to the university book store.

The common thread for almost all interviewees was the inability to respond to statements where the outcomes, especially related to career, were posed as certainties, for example:

1A) Pilot version: If I earn my undergraduate degree, I will be able to meet my financial goals

1B) Final version: If I earn my undergraduate degree, I will be more likely to meet my financial goals.

Some of the students suggested that career-related statements should be "nixed" and saved for upper level chemistry courses, when students are more certain of their goals and career plans. While social and self-evaluative outcome expectations were important to some students, especially those who were re-taking the course or going into medicine because they came from a family of doctors, most students cared little about the approval of family and friends. Lab related items required multiple revisions because most students did not necessarily "visualize the chemistry" while performing an experiment. This phrase had to be revised to "understand the chemistry". Three statements were excluded because they did not fit the idea of an outcome expectation during interviews and pilot testing:

a) If I am given an equation, I have no interest in its derivation (participants had no idea what the statement was asking and when clarified, most students offered a neutral rating about the item).
b) If I did not have to take exams in this chemistry course, I would have a better understanding of course material (students had mixed feelings about this statement as they understood the value of taking exams, but would like the option of not taking them because they were anxious test takers)
c) If I can use the correct equation or fact to obtain my answer, I can do well on quizzes/exams in this course (almost all students disagreed with this statement because they believe there was more to doing well on quizzes and exam than using the correct equation or fact).

Based on these revisions and removals, the final version of COES was developed. This version had 25 items hypothesized to measure expectations reflecting self-evaluative, social, physical reactions to personal behavior, expectations specific to academic performance and career and variations of behaviors and tasks in the CSEAS.

The final version of the COES are included in **Appendices J** and **K**.

A second round of interviews was conducted as part of the implementation phase of the instrument. The solicitation and interview processes were similar to those described in chapter 4 for the CSEAS. Eight students signed-up to participate in semi-structured think aloud interviews. Almost all students were biology majors on a pre-med track with an equal number of male and female participants. Compensation was a $10 gift card to the university book store and the interviews lasted 30-45 minutes.

<u>Data collection and participants</u>

The COES has been in administration since Fall 2013; the pilot study was conducted in Fall 2013 while data collection on the final version of the survey has been occurring since Spring

2014. The survey has been administered on paper since Fall 2013. The number of participants from the pilot and main administrations are shown (by course) in **Table 5.2** and **Table 5.3** respectively.

The pre-COES surveys for the pilot study were distributed and collected by teaching assistants in their respective discussions during the first week of class. In subsequent semesters, instructors distributed the pre-surveys on the first day of lecture with students returning them at the next lecture. Instructors were able to explain the purpose and the importance of the survey and encouraged students to complete it to the best of their ability. Students who did not complete their surveys or return their completed surveys within the first two weeks since start of classes were not included in data analyses; it was postulated that, in two weeks, these students had been sufficiently exposed to course material and the instructor for their responses to be influenced or biased. The post surveys were distributed a week or two before the start of final exam week. Surveys were distributed and collected in the same lecture or collected at the next lecture depending on the instructor's convenience. The surveys typically took 10-15 minutes to complete and students were given extra credit points for completing both surveys (COES and CSEAS).

The studies described in this chapter were conducted at a large, urban, research intensive public university in the Midwestern United States. Surveys were administered to students enrolled in preparatory chemistry, GC I, GC II and general chemistry for engineers; the descriptions of these courses are given in chapter 3.

Table 5.2. Participants (by course) for pilot administration of COES – Fall 2013

Fall 2013 (pilot)	Prep. Chem.	Gen. Chem. I	Gen. Chem. II	Gen. Chem. for engineers
Pre (N)	58	182	146	108
Post (N)	90	136	91	56

Table 5.3. Participants (by course) for main administration of COES – Spring 2014

Spring 2014 (main)	Prep. Chem.	Gen. Chem. I	Gen. Chem. II	Gen. Chem. for engineers
Pre (N)	348	209	146	93
Post (N)	311	182	115	60

Data analyses

Data were cleaned as described in chapter 3. Statements with conditions that were negatively worded had to be reverse scored before proceeding with any analyses. Although analyses were conducted on the pilot version as well, the results mandated major revisions in the pilot version. The focus in these analyses will be on data collected since Spring 2014 (first administration of the final version of COES).

Descriptive statistics were obtained for all items in the COES for assessments of univariate normality, skew, kurtosis and missing data.

<u>For data from GC I and GC II</u>

Factor analyses (EFA and CFA) was conducted to determine the most robust and interpretable factor structure. A mean score was calculated for each subscale ("factor") based on the raw responses to statements (items) that constituted the subscale. In an effort to obtain the most robust, meaningful factor structure from a fairly homogeneous dataset, EFA was not only conducted on pre- and post-GC I and GC II data respectively but also on a combination of GC II and GC I data. Among the datasets analyzed, the combination pre-GC II and post-GCI resulted in the most meaningful factor structure.

Comparative statistics were obtained (using GCI) as described in chapter 3. For independent sample t-tests, high vs. low performing groups (on final exam and in the course) were created based on z-scores for the raw data. Students with z-scores > 0 were categorized as the high-performing group and z-scores < 0 were the low-performing group. In the COES, a low mean score on a factor implied positive or high outcome expectations.

Reliability and validity were established using the measures detailed in chapter 3.

For data from preparatory chemistry and general chemistry for engineers

Cluster analyses were used to group the student responses from these courses. While these courses do not play an integral role in the development of a longitudinal model, they serve as two key courses that pave the way for students to be primed for enrollment in general chemistry or in their respective engineering fields. Thus, the analysis conducted here is the first step to examine the degree to which affective and cognitive meaning can be established in two courses comprising of highly heterogeneous groups of students.

Results and discussion

Descriptive statistics

The descriptive statistics for post-GC I, pre- GC II and the combined dataset are provided in **Table 5.4.**

Table 5.4. Demographic characteristics of a) post-GC I, b) pre-GC II and c) the combined dataset from Spring 2014

Post GC I				Pre GC II				Post-GC I + Pre-GC II		
Variable	N	%		*Variable*	N	%		*Variable*	N	%
Gender				Gender				Gender		
Male	81	44.5		Male	69	47.3		Male	150	45.7
Female	101	55.5		Female	77	52.7		Female	178	54.3
Acad. Level				Acad. Level				Acad. Level		
Freshman	34	18.7		Freshman	9	6.2		Freshman	43	13.1
Soph.	77	42.3		Soph.	54	37		Soph.	131	39.9
Junior	37	20.3		Junior	51	34.9		Junior	88	26.8
Senior	34	18.7		Senior	31	21.2		Senior	65	19.8

The mean ACT composite scores (not shown here) for post GC-I, pre-GC II and the combined dataset were 22.78, 23.97 and 23.29 respectively. The mean ACT Math scores were 22.57, 23.97 and 23.17 respectively while the mean ACT Sci-Re scores were 22.77, 24.09 and 23.33 respectively. The ACT scores and gender distributions reveal a fairly homogenous group of students in each course and in combination. In the combined dataset, the highest percentage of majors were in biology (24.7%), followed by undecided (13.1%) and biomedical sciences (11.6%). Over 95% of the students were in STEM fields, with the rest in non-STEM fields such as marketing and accounting. The percentage of missing item responses was 7% overall with the highest number being five missing responses for item 19. Although some individual items displayed skewness and kurtosis above recommended values, the resulting subscales or factors for the final model had values within range.

As different data sets were used for factor analysis and subsequent comparative statistics and reliability/validity testing, **Table 5.5** summarizes the number of students in the dataset, the course and the testing that was conducted using the corresponding dataset.

Table 5.5. Summary of datasets - number of students, course and time point for analyses

Analyses conducted	N	Course	Semester(s)
1. EFA (exploratory factor analyses)	312	GCI post + GC II pre	S14
2. CFA (confirmatory factor analyses)	152 - 377	GCI pre, post & GC II pre	F14, S15, S16, F16
3. Pre to post changes in COES subscales	368	GC I pre and post	S14 - S16
4. Differences in COES subscales by gender	368	GC I pre and post	S14 - S16
5. Reliability testing	315	GC I post + GC II pre	S14
6. Correlational analysis - subscales & performance indicators	354	GCI pre	S14 - S16
7. Correlational analysis - subscales & performance indicators	336	GCI post	S14 - S16
8. Differences in subscales (high vs. low performers on final exam)	368	GC I pre	S14 - S16
9. Differences in subscales (high vs. low performers in the course)	368	GC I pre	S14 - S16
10. Correlational analyses between pre-COES and pre-CSEAS subscale scores (validity testing)	367	GC I pre	S14 - S16
11. Correlational analyses between post-COES and post-CSEAS subscale scores (validity testing)	368	GC I post	S14 - S16

Factor analysis - Exploratory

Table 5.6 shows the factor structure of each dataset with descriptions of the results that led to using a combined dataset. The correlation matrix and item characteristics are only shown for the final factor structure. Items shaded in gray did not align between the two factors in that they showed loadings in one factor structure but did not load at all in the other structure. Similarly, items that loaded by themselves also did not align across factor structures.

Table 5.6. Component matrix: Six-factor solutions for post GC I and pre GC II respectively.

Item	Six factors - Post GC I					
	1	2	3	4	5	6
24		.730				
2						
1		.790				
22		.762				
12					.635	
17	.598					
21	.627					
15				.735		
18	.635					
14	.719					
9						
16	.752					
4			.782			
25			.714			
13						.522
19	.600					
11			.642			
23					.537	
10					.672	
5			.557			
7						
6				.607		
8						.768
3				.749		
20	.521					

Item	Six factors - Pre GC II					
	1	2	3	4	5	6
24	.806					
2	.753					
1	.750					
22	.747					
12	.632					
17	.599					
21	.572					
15						
18		.702				
14		.691				
9		.649				
16		.576				
4			-.755			
25			-.688			
13			.616			
19			.510			
11				.683		
23				.579		
10				-.562		
5					-.741	
7					.645	
6						
8						.650
3						
20						

Post GC I

The item responses in this dataset satisfied univariate normality and the correlation matrix satisfied KMO measures of sampling adequacy (.85 = good) indicating a dataset appropriate for factor analysis; using the eigenvalue > 1 condition resulted in six factors for the post GC I data. However, parallel analysis recommended a two-factor solution; in addition, correlating the factors and using an oblique rotation method such as promin resulted in one factor that did not make substantive sense. The factor correlation matrix indicated low degree of correlation among the factors. Thus, varimax was the preferred method of rotation and the factor structure shown in **Table 5.6** was retained.

Pre-GC II

The item responses in this dataset had some variables that displayed considerable skew and kurtosis. The matrix was deemed fair using the KMO statistic (.74). This suggests that, although a factor analysis may yield common factors, there were concerns about variables not loading on any factor or cross loading. As observed, the six-factor structure displays items that did not load or loaded in factors that were not as meaningful.

When comparing the two factor structures, there are some items that factor in post GC I but do not in pre-GC II and vice versa. When the number of extracted factors changes, previously 'unloaded' items have loadings. There were problematic items common to both structures which elicit removal or exclusion from analysis.

Thus, it was decided to combine both these datasets to facilitate a better dataset with a respectable sample size for factor analysis. The item means, standard deviations and inter-item correlation matrix for this analysis are presented in **Table 5.7**. On a 5-point scale, where 1=strongly agree and 5=strongly disagree, means ranged from 1.33 to 3.21. The correlation matrix does not show many correlations exceeding r=.70, thus indicating no problems with multicollinearity. Bartlett's test of sphericity was significant (χ^2 = 2230.4, p < 0.001), which indicated that the correlation matrix was not an identity matrix. The KMO statistics (.84) was good indicating that the matrix was appropriate for factor analysis.

Table 5.7. Correlation matrix, means and standard deviations for the chemistry outcome expectations scale (COES)

Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	Mean	SD
1	1.000																									1.56	.715
2	.391	1.000																								1.33	.547
3	.222	.291	1.000																							1.78	.770
4	-.014	.042	-.244	1.000																						2.46	1.054
5	-.102	-.032	-.246	.189	1.000																					3.02	.787
6	.252	.341	.250	.035	-.190	1.000																				1.71	.558
7	.032	.127	.221	-.121	-.203	.230	1.000																			2.21	.955
8	.043	.036	-.021	.195	.096	.137	-.032	1.000																		2.84	1.235
9	.405	.232	.136	-.020	-.095	.231	.148	-.009	1.000																	1.99	.769
10	.042	.122	.146	-.061	-.138	.079	.055	-.012	.133	1.000																2.24	.880
11	-.080	.099	-.053	.272	.163	.058	-.052	.011	-.049	-.038	1.000															2.52	1.780
12	.409	.307	.242	.013	-.204	.211	.084	-.008	.297	.204	.037	1.000														1.78	.749
13	.192	.150	.195	-.166	-.089	.199	.151	.099	.254	-.007	-.072	.290	1.000													2.10	.814
14	.202	.161	.112	.038	-.104	.213	.186	.045	.380	.121	-.044	.222	.196	1.000												2.10	.808
15	.256	.295	.551	-.069	-.129	.347	.278	.083	.224	.181	.021	.280	.230	.211	1.000											1.63	.644
16	.290	.319	.243	.101	-.104	.295	.192	.075	.497	.116	.059	.261	.178	.565	.301	1.000										1.93	.805
17	.329	.350	.185	.109	-.041	.366	.131	.066	.313	.089	.099	.329	.208	.348	.284	.529	1.000									1.69	.679
18	.205	.165	.185	-.076	-.058	.222	.193	.060	.386	.119	-.040	.201	.306	.345	.223	.379	.383	1.000								1.94	.746
19	.313	.227	.259	-.114	-.075	.235	.166	.000	.314	.116	-.088	.308	.302	.342	.327	.358	.360	.463	1.000							1.86	.712
20	.014	.052	.226	-.124	-.063	.059	.192	-.108	.308	.201	-.095	.044	.136	.197	.128	.212	.041	.317	.303	1.000						2.41	1.026
21	.238	.301	.177	.089	-.122	.263	.113	.067	.313	.189	.136	.296	.169	.342	.248	.492	.360	.338	.318	.231	1.000					1.73	.640
22	.577	.334	.201	.003	-.121	.248	.083	.100	.307	.121	-.070	.369	.278	.228	.233	.302	.387	.283	.396	.057	.337	1.000				1.61	.714
23	.112	.135	.189	-.055	-.074	.219	.267	-.003	.374	.218	.132	.194	.133	.148	.155	.241	.110	.242	.147	.341	.240	.110	1.000			2.16	.889
24	.543	.258	.197	.004	-.146	.278	.119	-.022	.301	.140	-.107	.335	.223	.261	.275	.303	.400	.290	.385	.133	.300	.706	.167	1.000		1.67	.745
25	-.013	.056	-.168	.494	.130	-.021	-.091	.147	-.026	-.119	.259	-.030	-.192	.023	-.105	.066	.054	-.087	-.129	-.228	.018	-.081	-.053	-.121	1.000	3.21	1.143

When the item-to-total scale correlations were examined, some of the items showed low or negative correlations, in addition to low square multiple correlations (item 5 = .194, item 8 = .139, item 10 = .157). When PCA analyses was conducted on this combined dataset using varimax rotation and the default eigenvalue > 1 criterion, a six-factor solution was obtained. Items 2 and 7 cross loaded, items 13 and 21 showed no loadings while items 5 and 8 either loaded by themselves or had negative loadings. The factor that loaded item 10 did not make substantive sense as this item loaded with two lab related items. Based on **Table 5.6**, these items were problematic in the individual factor structures as well. The scree plot, shown in **Figure 5.1** justified retention of five factors, while parallel analyses results in **Table 5.8** recommended retention of four factors; the cutoff point for parallel analysis was when the random order eigenvalue exceeded the actual data eigenvalue.

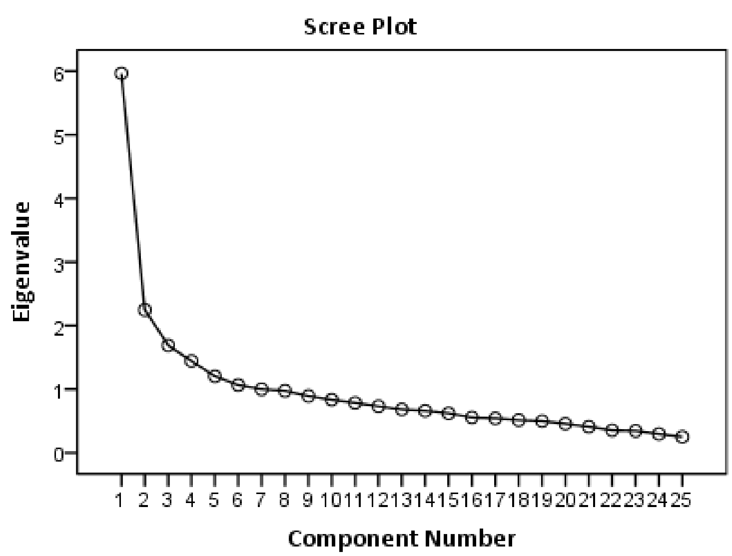


Figure 5.1. Scree plot showing eigenvalues for a five-factor solution

Table 5.8. Parallel analysis results showing eigenvalues for actual and random ordered data - COES

Eigenvalue #	Actual data eigenvalue	Random order eigenvalue	Percentile
1	5.96428	1.542	1.64483
2	2.24658	1.4554	1.53269
3	1.68895	1.3921	1.45284
4	1.44269	1.3368	1.39154
5	1.20394	1.2853	1.33602
6	1.06431	1.2400	1.28759

Based on these results, items 5, 8, 10, 13, 2, 7 and 21 were excluded from analyses to obtain a new structure. After several iterations, a five-factor solution was selected because it was the most conceptually interpretable factor structure. The total variance explained by the five factors was 58.2%. Variance explained by each factor was as follows: factor 1 = 15.0%, factor 2 = 13.9%, factor 3 = 10.8%, factor 4 = 9.9% and factor 5 = 8.7%.

The factor loadings from the rotated component matrix are shown in **Table 5.9**. Factor names and items in each factor are shown in the table. Items with asterisks were reverse scored before analyses. This factor structure was tested using CFA; the final model for the COES is represented in the path diagram. This model accounts for any modifications resulting from CFA to improve model fit.

Each factor was named based on the tasks producing the outcomes or the outcomes themselves. Subscales that involved outcomes related to career were named 'outcome expectations related to career goals/planning", items that were related to task based outcomes were named according to the nature of the tasks being performed. The subscale that was named outcome expectations related to performance based tasks was done so according to the meaning of performance based learning, which represents a set of strategies for acquisition and application of knowledge, skills and work habits through the performance of tasks that are meaningful and engaging; similarly, the factor relating outcome expectations to understanding chemistry was named based on whether the condition or outcome of the statement involved a task that necessitated understanding chemistry either on a fundamental level or applying higher order thinking for abstract tasks such as ‘changing ideas about how the physical world works’.

Table 5.9. Rotated component matrix for five-factor solution: COES (post GC II + pre-GC I; N = 312). Items with asterisks were reverse scored.

Factor & Item	Item	Five factors - Pre GC II + Post GC I				
		1	2	3	4	5
Factor 1	**Outcome expectations related to performance based tasks**					
1	If I work hard enough, I will be more likely to pass this course	.815				
22	If I do everything possible (for example, review class notes...), I will do well in this course	.804				
24	If I do everything possible (for example, review class notes...), be prepared for quizzes/exams in this course	.771				
12	If I obtain a good grade in this course, I will have a better chance of achieving career goals	.582				
Factor 2	**Outcome expectations related to understanding chemistry**					
14	If I learn chem, I expect to change some of my ideas about how the phys.world works		.780			
16	If I can relate chem to situations in everyday life, I expect to learn it better		.731			
18	If I understand a fundamental concept, I can solve homework/exam problems on that concept		.591			
17	If I figure out what I did wrong on my exam, I will improve my understanding of course material for the next exam		.556			
19	If I can follow my instructor in lecture, I expect to do better in this course		.506			
Factor 3	**Outcome expectations related to career planning and choices**					
15	If I succeed at getting my intended degree, I will be more likely to achieve my career goals			.776		
3	If I graduate with my current major, I will be more likely to get a well paying job			.746		
6	If I know my interests and abilities, then I will make better career decisions			.534		
Factor 4	**Outcome expectations related to learner based tasks**					
4*	If all I do is memorize the solution to any problem, I will be successful in this course				.743	
25*	If I can remember the solution to a problem and know where to put numbers, do well on quizzes/exams				.738	
11*	If I don't understand the concepts in this course, I can pass (with at least a C)				.668	
Factor 5	**Outcome expectations related to success in lab**					
23	If I understand the principles behind the experiments, I will be more likely to succeed in laboratory					.830
20	If I finish my experiment and while in lab, figure out what my data means, I expect to do well in laboratory					.659
9	If I try and understand the chemistry while performing an experiment, I will do well in laboratory					.501

Factor analyses – Confirmatory

The factor structure detailed in **Table 5.9** was imposed on multiple datasets to check model fit at each time point during the two-semester gateway sequence. Imposing one factor structure was necessary to fix items so reasonable subscore comparisons could be facilitated. In addition, obtaining a pure, meaningful factor structure (with sensibly grouped items) was the first step towards making meaningful measurements longitudinally.

Although the SAS code (shown in **appendix A**) requested for outliers, these were not excluded arbitrarily. Fit indices were checked with and without exclusion of outliers, along with other generated output, to make decisions about model fit. While the above data provided a model with reasonable fit, indices improved considerably when item 12 was moved from factor 1 (expectations related to performance based tasks) to factor 3 (expectations related to career outcomes). The results shown in **Table 5.10** are for models that have incorporated this change. Not all fit indices are shown in this table. Detailed descriptions of the indices shown have been included in chapter 3.

Table 5.10. Goodness-of-fit indicators for COES models tested at three time points during AY14-15.

Model tested on:	N	$\chi2$	df	$\chi2$ / df	SRMR	CFI	RMSEA	RMSEA CI
F14 + S15 pre -GC I	377	277.06	125	2.21	.057	.922	.057	0.0479 - 0.0659
F15 + S16 pre-GC I	369	252.25	125	2.01	.057	.941	.053	0.0436 - 0.0623
F14 + S15 post-GC I	289	263.71	125	2.10	.056	.941	.060	0.0516 - 0.0725
F15 + S16 post-GC I	287	289.74	125	2.31	.069	.922	.070	0.0631 - 0.0832
F14 + S15 pre-GC II	259	273.29	125	2.18	.060	.905	.068	0.0568 - 0.0786
F15 + S16 pre-GC II	287	288.90	125	2.31	.066	.921	.068	0.0575 - 0.0780
F16 pre-GC I	226	249.31	125	1.99	.058	.927	.067	0.0544 - 0.0785
F16 pre-GC II	152	192.15	125	1.54	.066	.936	.060	0.0422 - 0.0759

As shown in the **Table 5.10**, the indices display reasonable – good model fit at each time point as indicated by SRMR values less than the recommended value of .08, CFI values > .90 and RMSEA values either at or less than the recommended value of .06. Although chi-square values were significant for all test datasets, this was to be expected with N > 200 in almost all datasets. Despite combining datasets, the clarity in factor structure and model fit are indications that students are responding to items using similar context associations. However, ongoing factor structure and model fit examinations are necessary to ensure the stability of the model as some of the fit indices for post data sets are above recommended values. The path diagram for the finalized COES model is shown in **Figure 5.2.**

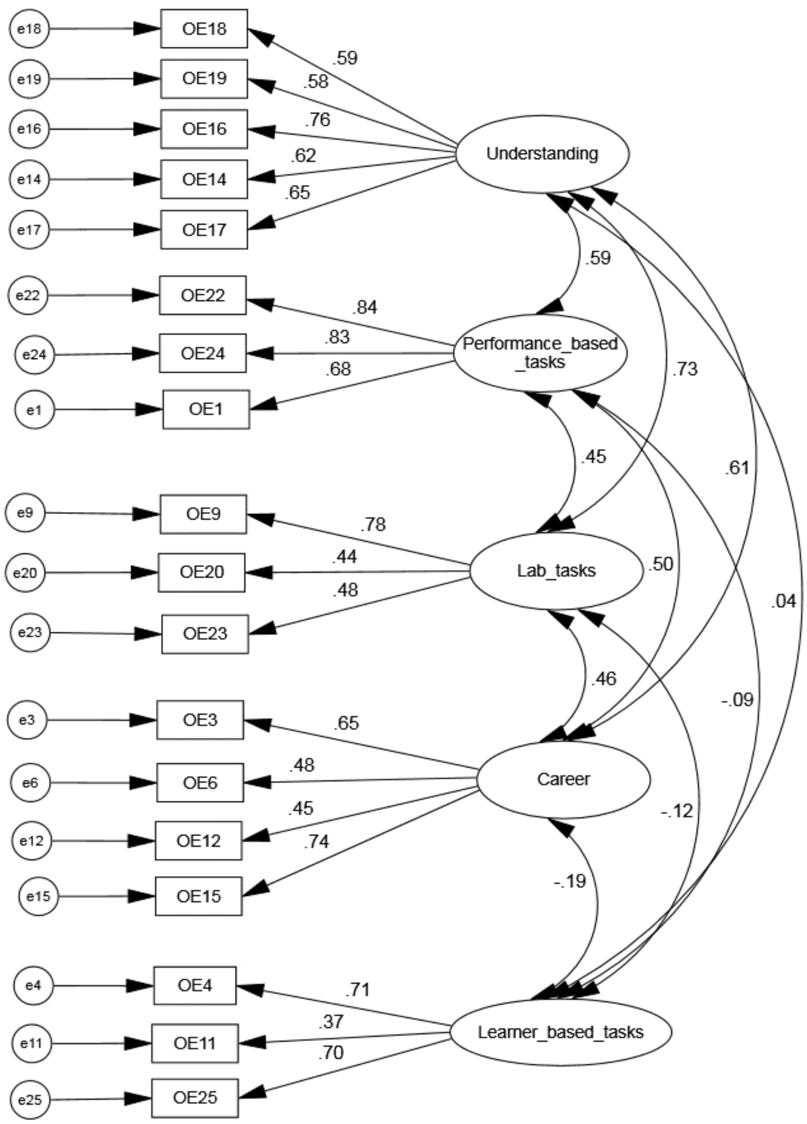


Figure 5.2. Standardized coefficients for the final, refined five-factor model of the chemistry outcome expectations scale. All coefficients are significant at p <0.01.

Comparative statistics

Pre-post score changes on each subscale (factor) were examined using the dataset indicated earlier in **Table 5.5**. Pre- and post-GC I samples was used as the test datasets for these analyses.

Scores for each multi-item subscale were obtained by calculating mean response scores for the items constituting the subscale (after reverse scoring items with negatively worded conditions). These results are shown in **Table 5.11**.

Table 5.11. COES scores showing pre to post changes for GC I (N = 368, S14-S16)

Factor	Avg. prettest scores	Avg. posttest scores	t	*p*	Effect size
1. Expectations related to performance based tasks	1.35	1.76	-10.822	<0.0001	0.720
2. Expectations related to understanding chem.	1.73	1.92	-5.786	<0.0001	0.383
3. Expectations related to career planning	1.57	1.73	-5.387	<0.0001	0.340
4. Expectations related to learner based tasks	2.71	2.75	-0.693	0.489	0.040
5. Expectations related to success in lab	1.84	1.89	-1.012	0.312	0.071

Aside from expectations related to learner based tasks and success in lab, the other subscales had significantly higher posttest scores, indicating less positive expectations related to the outcomes in each subscale. Expectations related to performance based tasks showed the greatest numerical increase (nearly a full standard deviation), indicating that students had the least positive (lowest) expectations about the outcomes related to these tasks. Expectations related to understanding chemistry also showed a significant increase in subscale score, indicating that students had less positive expectations about understanding chemistry; similarly, expectations related to career planning also showed a significant increase in average subscale scores, implying that students had lower expectations related to career planning/goals at the end of the semester.

It is quite possible that, despite making changes to the statements to reflect a future chemistry course, students were responding to post survey items by reflecting on everything they had done during the semester to judge their performance outcome expectations. Given that some students stated during interviews that "doing everything possible for the course wouldn't necessarily ensure doing well in the course", it is possible that such beliefs have may resulted in negative expectations at the end of the course. Expectations related to career outcomes, while still significantly less positive posttest, did not show a drastic increase quantitatively as indicated by the effect size. This could indicate that although career expectations form and evolve through the course of GC I, they are distal outcomes which probably get established better as students transition through chemistry courses necessary for their major. The lack of a significant difference for expectations related to learner based tasks could reflect students' global vs. domain specific beliefs that understanding concepts rather than mere memorization is essential to succeed not just in chemistry but in other courses too. For expectations related to success in lab, it is possible that students viewed lab as an independent entity and not as a complement to the material covered in lecture; thus, they could have an indifferent view towards the tasks integral to success in lab. At the same time, if students working in groups were vicariously successful in lab due to their lab partners, they could have a falsified sense of positive outcome expectations. While these are changes observed overall, differences could be manifested in other ways when observed by subgroups.

The results of examining the factors by gender subgrouping are shown in **Table 5.12**. For ease of interpretation, only significant t-test results are displayed.

Table 5.12. Results showing differences in COES pre-subscales based on gender (N_{males} = 152, $N_{females}$ = 215) – GC I (S14-S16)

Factor	Males	Females	t	*p*	Effect size, d_{Cohen}
1. Expectations related to performance based tasks	1.28	1.39	-2.421	0.016	0.251
4. Expectations related to learner based tasks	2.87	2.60	3.402	0.001	-0.361
5. Expectations related to success in lab	1.73	1.92	-2.574	0.010	0.264

Subscale differences based on gender reveal that females have less positive outcome expectations than males with regards to performance based tasks and success in lab but display more positive outcome expectations than males with regards to learner based tasks. To examine the impact of prior ability on these differences, 117 males and 182 females were examined by ability level; the placement tests (TP and ACT) were used as proxies for prior ability. Male and female students differed significantly in their performance on the TP Chemistry placement test (t (227) = 2.526, p = .012) and TP total (t (297) = 2.630, p = .011). Based on these differences, the subscales shown in **Table 5.12** were examined for differences between male and female students of high vs. low ability respectively in both TP chemistry and TP overall. When examined by ability in TP chemistry, no significant differences were observed between high ability males and females on any subscale shown in **Table 5.12**.

However, when low ability students were examined, female students showed more positive expectations than males about performance outcomes related to learner based tasks such as memorization of material without understanding concepts. In addition, female students showed less positive expectations with regards to success in lab and performance based tasks. These results are summarized in **Table 5.13**.

Table 5.13. Results showing gender differences within subscales for low ability students (based on TP chemistry); (N_{males} = 62, $N_{females}$ = 126) – GC I pre-COES, S14-S16

Factor	Males	Females	t	p	Effect size, d_{Cohen}
1. Expectations related to performance based tasks	1.31	1.46	-2.072	.040	.312
4. Expectations related to learner based tasks	3.02	2.63	3.251	.002	-.514
5. Expectations related to success in lab	1.72	1.98	-2.679	.008	.380

When TP total was used as the indicator for ability, no significant differences were observed between high ability males and females on either subscale. However, low ability females showed more positive expectations than males with regards to learner based tasks; they also demonstrated more negative expectations than males with respect to success in lab.

Reliability

Factor correlations and factor alpha coefficients were calculated for the model confirmed by CFA. These results are shown in **Table 5.14**.

Table 5.14. Factor correlations and Cronbach alpha coefficients for the COES (N = 315, S14 GC II pre + GCI post) **Correlation is significant at the 0.01 level. *Correlation is significant at the 0.05 level.

Factor	Mean	Std. dev.	1	2	3	4	5	Cronbach's alpha (α)
1. Expectations related to performance based tasks (n=3)	1.62	.62	1.000					0.814
2. Expectations related to understanding chem. (n=5)	1.90	.54	.492**	1.000				0.767
3. Expectations related to career planning (n=4)	1.72	.48	.459**	.480**	1.000			0.636
4. Expectations related to learner based tasks (n=3)	2.73	.78	-.072	.005	-.084	1.000		0.607
5. Expectations related to success in lab (n=3)	2.20	.67	.278**	.460**	.331**	-.118*	1.000	0.602
Total scale (n=18)								0.770

As the factors were expected to correlate a certain extent, moderate correlations were expected between expectations related to understanding chemistry and expectations related to performance based tasks as a huge component of performance based learning involves tasks that require a strong understanding of material. On the other hand, weak correlations were expected between expectations related to learner based tasks and understanding chemistry as the tasks and outcomes for learner based subscale did not necessitate much understanding of material. Moderate correlations were also expected between expectations related to success in lab and performance based tasks due to lab success being one of the tasks that dictated high performance in the course. It was expected that expectations related to career would display weak correlations with some of the subscales; as this dataset contained GC II as well, perhaps the decisions about career goals were better established and could result in some moderate correlations.

Based on the results in **Table 5.14**, moderate and significant correlations were observed between expectations related to understanding chemistry and performance based tasks as expected; similar strength of correlations was also observed between expectations related to career planning and performance based tasks, between expectations related to career planning and understanding chemistry and expectations related to success in lab and understanding chemistry. The negative values indicate an inverse relationship between expectations related to learner based tasks and performance tasks; in addition, as expected, extremely weak and non-significant correlations were observed between expectations related to learner based tasks and understanding chemistry. Negative and non-significant correlations were also observed between expectations related to learner based tasks and career planning, indicating that students who had positive expectations about learner based tasks such as memorizing material to pass the course had negative expectations about career planning and setting career goals. The significantly negative correlation between expectations related to learner based tasks and success in lab indicates that students with positive expectations about learner based tasks such as memorizing material had low expectations about success in lab. Given the reasonable degree of complexity in lab related tasks, starting with preparation before the experiment until after completion, it would not seem unreasonable that students with highly positive expectations about learner based tasks would have less positive expectations about success in lab, where a level of thinking above memorizing material might be warranted.

Reliability estimates were calculated using the same dataset that was used in factor analysis (S14 GCII pre + GCI post, N=315); these estimates ranged from .60 to .81 with a total scale Cronbach's alpha equal to .770. When the scale with 25 items was evaluated, Cronbach's alpha was equal to .776. Although the scale displayed some low item-total correlations, when the item was examined within its subscale, the low item-total correlation was no longer prevalent. When tested on a GC I pre dataset, the estimate was .745 and with a post dataset, it was .801. As some of the subscales displayed low reliability estimates and reliability for an instrument in production was required to be above 0.80 (Nunnally, 1978), this structure was to be monitored and tested consistently for appropriate model fit.

Validity

Construct validity was established using exploratory and confirmatory factor analyses. The presence of distinct factors measuring different aspects of outcome expectations suggest that the survey captures dimensions of this construct that would otherwise be immeasurable using performance indicators alone.

Predictive validity (a form of criterion-related validity) was evaluated by correlating mean subscale scores to placement test scores, final exam and course performance percentages. These correlations are displayed in **Table 5.15**. Expectations related to career outcomes did not show a significant correlation with any performance indicator; as these were pre-COES subscales, it is possible that students have not thought about outcomes or expectations related to career goals for this subscale to have a significant relationship with measures of cognitive ability or exam/course performance.

Table 5.15. Correlations between pre- COES subscale scores and performance indicators in GC I (N = 354, S14-S16). **Correlations significant at 0.01 level; *Correlations significance at 0.05 level.

Factor pre GC I	ACT Comp.	ACT Math	ACT-SciRe	Final Exam %	Course %
1. Expectations related to performance based tasks	-.108*	-.079	-.118*	-.154**	.180**
2. Expectations related to understanding chem.	-.116*	-.093	-.101	-.156**	-.154**
4. Expectations related to learner based tasks	-.239**	-.158**	-.150*	-.238**	-.213**
5. Expectations related to success in lab	-.069	-.065	-.116*	-0.09	-.170**

The results in **Table 5.15** support the hypothesis from SCCT's model of performance that a positive relationship exists between positive outcome expectations and career/academic performance. The correlations are negative because in the COES, a high score implies negative expectations (1=strongly agree and 5=strongly disagree). Expectations related to success in lab demonstrate a significant correlation to course performance but not to performance on the final exam; given the lab component's influence on the course grade as opposed to the final exam could explain this significant correlation. Weak to moderate correlations between the subscales and performance measures indicate that the COES is not just another measure of academic ability or achievement.

When these correlations were repeated using the post subscale scores, they became slightly stronger as the outcome expectations were assessed at a time point close to the final exam; in addition, as Bandura suggests, it is likely that the post expectations were influenced partially by self-efficacy beliefs as the outcomes were closely linked to the quality of one's performance (Lent et al., 1994; Bandura, 1986). Expectations related to career planning become significant with some performance measures. Despite the increased strength in correlations between the post-subscale scores and performance indicators, the performance and persistence models developed in this project utilized pre measures because a) the interest was in examining affective measures of students entering the course as opposed to how the course impacted these measures and b) the pre measures allowed models to account for students coming in with varied ability levels as they would be predicted to do better or worse accordingly. While using post measures would increase variance explained by the models, there would be no way of knowing if this was because of an increased affective measure due to increased ability or some other reason beyond performance. Results are shown in **Table 5.16**.

Table 5.16. Correlations between post- COES subscale scores and performance indicators in GC I (N = 336, S14-S16). **. Correlations significant at 0.01 level; *. Correlations significance at 0.05 level.

Factor	ACT Comp.	ACT Math	ACT-SciRe	Final Exam %	Course %
1. Expectations related to performance based tasks	-.085	-.057	-.069	-.242**	-.238**
2. Expectations related to understanding chem.	-.098*	-.051	-.062	-.216**	-.221**
3. Expectations related to career planning	-.035	.002	-.044	-.112*	-.129**
4. Expectations related to learner based tasks	-.227**	-.170**	-.189**	-.128**	-.130**
5. Expectations related to success in lab	-.106*	-.064	-.122*	-.158**	-.158**

When differences were examined by high vs. low performing student groups based on final exam performance, significant differences were observed between both groups on each subscale. High performers consistently showed lower average scores (more positive outcome expectations) than the low performing group. These results are shown in **Table 5.17**.

Table 5.17. Results showing differences in COES pre-subscales based on high vs. low performing groups on the final exam ($N_{\text{low pefomers}}$ = 173, $N_{\text{high performers}}$ = 195) – GC I (S14-S16)

Factor	Low performers	High performers	t	*p*	Effect size
1. Expectations related to performance based tasks	1.44	1.26	3.837	<0.0001	-0.407
2. Expectations related to understanding chem.	1.82	1.66	3.532	<0.0001	-0.369
3. Expectations related to career planning	1.62	1.52	2.039	0.010	-0.213
4. Expectations related to learner based tasks	2.82	2.61	2.591	0.042	-0.271
5. Expectations related to success in lab	1.95	1.75	2.747	0.006	-0.287

When this process was repeated using performance in the course, significant differences emerged between both groups for all subscales, except expectations related to career planning. High performers consistently showed lower average scores (more positive outcome expectations) than the low performing group. It is possible that neither high nor low performing students have fully mapped out their expectations related to career tasks as GC I might be fairly premature for students to start thinking about fulfilling financial and career goals. Given that career outcome expectations are environmental contingencies, students are less likely to have any expectations with regards to outcomes that are beyond their control regardless of their performance; for instance, during interviews, several students were cautious about having any expectations at all with regards to job related outcomes, especially when the outcomes were influenced by external factors e.g. "graduating with their intended majors would not necessarily earn them well-paying jobs as the job market would influence the situation". These correlations are shown in **Table 5.18**.

Table 5.18. Results showing differences in COES pre-subscales based on high vs. low performing groups in the course ($N_{\text{low performers}}$ = 168, $N_{\text{high performers}}$ = 200) - GC I (S14-S16)

Factor	Low performers	High performers	t	*p*	Effect size
1. Expectations related to performance based tasks	1.44	1.27	3.662	<0.0001	-0.393
2. Expectations related to understanding chem.	1.79	1.68	2.250	0.025	-0.235
3. Expectations related to career planning	1.61	1.54	1.450	0.148	-0.152
4. Expectations related to learner based tasks	2.81	2.63	2.266	0.025	-0.236
5. Expectations related to success in lab	1.95	1.75	2.835	0.005	-0.297

Validity of the subscales was also tested by examining the degree to which the subscales measuring different aspects of outcome expectations were related or unrelated to other operationalized measures. Thus, average subscale scores were correlated with students' certainty of persisting in their majors (1= very certain to 5=very uncertain). This was an item ('How certain are you of persisting in your intended major?') that constituted the supplemental items in the CSEAS survey. It was expected that students with positive expectations related to some of the subscales, particularly career goals, would display significant and moderate correlations with their certainty of persisting in a major. For GC I pre-COES subscale scores (S14-F14)), there was a significant, positive correlation between outcome expectations related to career planning and students' certainty of persisting in their majors (N = 283, *r*=.185). Students with more positive outcome expectations related to careers were more certain of persisting in their majors.

When COES subscale scores were correlated with the four items from CSEQ, each measuring how often students a) discussed course information with their instructor (grades, possible make-up work, assignments) b) applied material learned in class to other areas (job/internship, other courses, interactions with others) c) memorized formulas, definitions, technical terms and concepts and d) discussed career plans and ambitions with anyone (advising staff, faculty members, family members) on a scale of 1=never, 4=very often and 5=not sure), it was expected that expectations related to learner based tasks would correlate to a moderately significant degree to students' memorizing formulas as learner based tasks involved memorization as well; it was also expected that expectations related to performance based tasks and understanding chemistry would correlate with applying material to other areas quite often.

The results indicated the following significant correlations: As expected, students who had more positive expectations related to performance based tasks and understanding chemistry applied material they had learned in class to other areas (r=.240 and r=.272 respectively) more often. Students with positive outcome expectations about their success in lab rarely discussed course information with their instructors (r= -.269). Those with positive outcome expectations related to learner based tasks memorized formulas, definition and concepts more often (r=.116). Lastly, students with positive outcome expectations related to understanding chemistry and career planning discussed their career plans and ambitions more often (r=.118 and r=.236 respectively). These results suggest that the outcome expectation subscales are relating meaningfully and as expected with items or scales that share commonality with tasks in some of these subscales.

In addition, convergent validity was examined by correlating subscales from the COES with those from the CSEAS. Results were evaluated for pre- and post-subscale scores for GC I.

The results for correlations obtained using pre-COES and pre-CSEAS subscales are shown in **Table 5.19.**

Table 5.19. Correlations between pre-COES and pre-CSEAS subscale scores in GC I (N = 367, S14-S16). **. Correlations significant at 0.01 level; *. Correlations significance at 0.05 level.

Factor	SE - assessment & exam prep	SE - interpersonal	SE- strategies	SE-low order tasks	SE-high order tasks	SE-everyday tasks involving chem.
1. Expectations related to performance based tasks	.409**	.275**	.265*	.141**	.129*	.247**
2. Expectations related to understanding chem.	.301**	.207**	.132*	.161**	.170**	.165**
3. Expectations related to career planning	.289**	.204**	.084	.025	.073	.130*
4. Expectations related to learner based tasks	-.021	.014	-.015	-.045	-.113*	.005
5. Expectations related to success in lab	.294**	.236**	.120*	.168**	.139**	.085

Given the complete and partial relationship between self-efficacy and outcome expectations described earlier in this chapter, correlations were expected between certain COES and CSEAS subscales such as self-efficacy in assessment vs. expectations related to performance based tasks, expectations related to understanding chemistry and to a small extent expectations related to learner based tasks. According to Lent et al., (1994), outcome expectations would relate strongly to self-efficacy, especially when outcomes are closely tied to the quality of one's performance. Lent et al., observed a significant correlation (r=.49) between self-efficacy and outcome expectations.

This is in alignment with the correlation (r=.409) between efficacy beliefs related to assessment and expectations related to performance outcomes. It was also expected that subscales sharing commonalities in tasks would correlate strongly. Thus, it was expected that expectations related to performance based tasks would correlate strongly with self-efficacy related to assessment and exam preparation; in addition, expectations related to understanding chemistry would correlate moderately with self-efficacy related to higher order tasks or applying chemistry strategies. Efficacy related to interpersonal tasks was expected to correlate weakly with some

COES subscales, but a strong correlation was expected between interpersonal self-efficacy and expectations related to understanding chemistry; the outcome of understanding chemistry could perhaps correlate to a certain degree with students' confidence about asking questions or interacting with the instructor. Career related outcome expectations were not expected to correlate strongly to any efficacy subscales as these were pre-subscales and expectations related to career might have just started developing.

The correlations between expectations related to understanding chemistry and all CSEAS subscales suggest a reasonable degree of commonality between the tasks involved in understanding chemistry and the self-efficacy subscales. Less expected were correlations between subscales in the CSEAS and expectations related to career planning. Although CSEAS subscales were expected to correlate weakly with expectations related to learner based tasks, the only significant correlation was between self-efficacy related to higher order tasks and expectations related to learner based tasks. The results for correlations between post subscale scores are shown in **Table 5.20**.

Table 5.20. Correlations between post-COES and post-CSEAS subscale scores in GC I (N = 368, S14-S16). **. Correlations significant at 0.01 level; *. Correlations significance at 0.05 level.

Factor	SE - assessment & exam prep	SE - interpersonal	SE- strategies	SE-low order tasks	SE-high order tasks	SE-everyday tasks involving chem.
1. Expectations related to performance based tasks	.476**	.428**	.341**	.130*	.223*	.210**
2. Expectations related to understanding chem.	.354**	.332**	.243*	.157**	.194**	.219**
3. Expectations related to career planning	.125*	.139**	.127*	.084	.106*	.146**
4. Expectations related to learner based tasks	-.091	-.072	.012	.094	-.027	.137**
5. Expectations related to success in lab	.223**	.216**	.246**	.118**	.226**	.239**

The posttest scores from both surveys show stronger correlations than pretest measures possibly because the outcomes (exam or course performance) are more closely tied to the quality of a student's performance either on a high-stakes assessment measure like the final exam or lab practical or other components that contribute towards their grade in the course. The moderate correlations between subscales from both constructs support the contention that SE and OE are distinctly different yet related constructs.

As a final check of validity, results from student interviews were examined for item groupings. Some of the common groupings resulting from these interviews were as follows:

a) Items 22 and 24 (expectations related to performance based tasks) were selected to be grouped together by almost all students. This was similar to the factor structure (**Table 5.9**), in which items 22 and 24 always grouped together due to the similarity in tasks, despite different outcomes.

b) Items 1,3,5,6,7,8,12 and 15 were grouped as "future plans and career". In one case, item 1 was grouped with other items that had performance outcomes such as items 4,9,11,13,16,17,18,19 and 20. While these items were categorized as career, not all of them were part of the corresponding subscale in the factor structure because of being standalone items or items which consistently cross loaded and warranted removal (items 5,8,7). However, the fact that items 5 and 7 were measuring career outcomes and categorized appropriately by students is an indication that items were being interpreted in similar contexts.

c) Items 13,20 and 23 were lab related items. While item 13 was dropped from the final model, items 20 and 23 were measuring components of success in lab.

d) 2,6 and 10 were labeled "self". While these items (purported to measure self-evaluative, physical and social outcome expectations) did not consistently load or form substantively meaningful groupings with other items, two of them (2 and 10) were excluded from the final factor structure. However, given that students recognize items as possibly measuring some component of "self", demonstrates the similar contextual associations made with regards to these items.

e) 5 and 8 were called "major".

In general, students came up with categories like careers/jobs, learning and performance.

Concerning the items that were removed to produce the final COES model, students made the following comments during interviews:

A) JF, a male student in computer science said the following about item 10 – 'If I make a good career decision, then my family and friends will approve of me'.

"I do not like this item as a career based decision. It is equivalent to 'going to church' so others see you as a religious person; it has no bearing on the thought process for career decisions. If this were a non-career based decision, it would make more sense because making poor life decisions would invoke approval or disapproval and impact others around you".

He also mentioned that item 2 – "If I do well/get a good grade in this course, I will be proud of myself"- would elicit a neutral response because "it is expected of me that I will do well". The student was focused more on the condition than the outcome because he mentioned that "the converse condition would probably elicit a different response from me".

B) TS, male student with a major in biology intending to go into pre-med stated about item 8 – 'If I am unable to pass this course, I will be more likely to change my major'.

"Throw out item 8. It is difficult to judge if major will change based on only one course. The item would be a lot more useful at higher level chemistry courses"

The student also stated, with regards to item 7 – 'If I earn my undergraduate degree, I will be more likely to meet my financial goals'.

"I do not necessarily agree with this statement because having a degree does not always relate to financial goals. The condition and outcome are not related. You have to do what makes you happy. It is a perception of people that earning a degree means you will make more money and be happy"

When comparing factor structures, items grouped by students aligned fairly well with factors generated quantitatively. Given the variability of students and the fact that the context was not explicitly stated for the items, it was not an easy task to determine context association, especially for the classwide data. It was expected that either the items would factor (as they were preliminarily) or they would not. As the items factored into meaningful subscales, similar to the item groups students had created, it was either because students were using the same or similar context association or the association was much less important. In this case, it seems the context used by the students (and the commonality of this) was fairly similar, thus offering some support for similarity in the classwide data as well.

Cluster analyses

The cluster structures resulting from data collected in preparatory chemistry and general chemistry for engineers are shown. As stated earlier, only the cluster structures are displayed here (with information about items that were excluded) in an effort to examine the utility of the COES survey. Items with asterisks were reverse scored before proceeding with analyses.

<u>Preparatory chemistry</u>

The dataset used for obtaining the cluster structure consisted of 628 students, out of whom 58.6% were females and 41.4% were males. 47.4% were freshmen, 36.4% were sophomores, 10.5% were juniors and 5.4% were seniors. The average ACT composite, math and sci-re scores were 22.3, 22.2 and 22.5 respectively. Based on a range of cluster solutions, items 2,5,8,10 and

11 were excluded to obtain the most meaningful four-cluster solution, shown in **Table 5.21**. The factor structure for preparatory chemistry data shows some factors common to the final factor structure (**Table 5.9**). The cluster describing expectations related to learner based tasks contains two items common to the subscale in the final factor structure. The subscale and cluster related to career planning and understanding also consist of similar items. In the case of preparatory chemistry though, there was no discrimination between the clusters corresponding to performance based tasks and other items containing outcomes related to course performance as there was in the factor structure for general chemistry. Items were grouped together based on their outcomes being related to course performance. Given the absence of a laboratory component in preparatory chemistry, it is possible that students perceive items related to course performance as being in the same category regardless of whether the performance is in the course or in laboratory. There were also two items that clustered without any meaning to the cluster.

<u>General chemistry for engineers</u>

The dataset used for obtaining the cluster structure consisted of 839 students, out of whom 57.8% were females and 42.2% were males. 47.3% were freshmen, 36.8% were sophomores, 9.8% were juniors and 6.0% were seniors. The average ACT composite, math and sci-re scores were 22.4, 22.5 and 24.0 respectively. Based on a range of cluster solutions, items 5,8 and 10 were excluded to obtain the most meaningful five-cluster solution, shown in **Table 5.22**. The data resulting from these students followed a similar pattern to preparatory chemistry in that there was no discrimination among items whose outcomes were related to course performance. This cluster structure displayed a cluster, containing the same items as the structure in **Table 5.9**, for expectations related to lab success. Data from general chemistry for engineers showed more meaningful clusters. However, several items in this structure were also grouped into one large cluster related to course performance.

As quite a few students in engineering are also in the workforce, the interests and goals of these students could be well crystallized to make assessments about their careers and financial goals. In addition, as the context of a 'future chemistry course' does not apply to these students given the terminal nature of this course, it is possible that students made retroactive associations when responding to some of the statements in the post survey.

Table 5.21. Four-cluster solution for preparatory chemistry students – COES (N = 628)

Cluster & Item	Item
Cluster 1	**Outcome expectations related to course performance**
1	If I work hard enough, I will be more likely to pass this course
17	If I figure out what I did wrong on my exam, I will improve my understanding of course material for the next exam
18	If I understand a fundamental concept, I can solve homework/exam problems on that concept
19	If I can follow my instructor in lecture, I expect to do better in this course
20	If I finish my experiment and while in lab, figure out what my data means, I expect to do well in laboratory
21	If I can explain a problem or concept to a classmate, I will understand the material better
22	If I do everything possible (for example, review class notes…), I will do well in this course
23	If I understand the principles behind the experiments, I will be more likely to succeed in laboratory
24	If I do everything possible (for example, review class notes…), be prepared for quizzes/exams in this course
Cluster 2	**Outcome expectations related to career planning**
3	If I graduate with my current major, I will be more likely to get a well paying job
7	If I earn my undergraduate degree, I will be more likely to meet my financial goals
15	If I succeed at getting my intended degree, I will be more likely to achieve my career goals
Cluster 3	**Outcome expectations related to learner based tasks**
4 *	If all I do is memorize the solution to any problem, I will be successful in this course
25 *	If I can remember the solution to a problem and know where to put numbers, do well on quizzes/exams
Cluster 4	**Outcome expectations related to understanding chemistry**
9	If I try and understand the chemistry while performing an experiment, I will do well in laboratory
14	If I learn chem, I expect to change some of my ideas about how the phys.world works
16	If I can relate chem to situations in everyday life, I expect to learn it better
Factors whose items do not make substantive sense	
12	If I obtain a good grade in this course, I will have a better chance of achieving my career goals
13	If I am able to follow the procedure to perform an experiment, I can understand what is happening in a future chemistry laboratory course

Table 5.22. Four-cluster solution for general chemistry for engineers - COES (N = 839)

Cluster & Item	Item
Cluster 1	**Outcome expectations related to course performance**
1	If I work hard enough, I will be more likely to pass this course
2	If I do wel/ get a good grade in this course, I will be proud of myself
12	If I obtain a good grade in this course, I will have a better chance of achieving my career goals
13	If I am able to follow the procedure to perform an experiment, I can understand what is happening in a future chemistry laboratory course
14	If I learn chem, I expect to change some of my ideas about how the phys.world works
16	If I can relate chem to situations in everyday life, I expect to learn it better
17	If I figure out what I did wrong on my exam, I will improve my understanding of course material for the next exam
19	If I can follow my instructor in lecture, I expect to do better in this course
22	If I do everything possible (for example, review class notes...), I will do well in this course
24	If I do everything possible (for example, review class notes...), be prepared for quizzes/exams in this course
Cluster 2	**Outcome expectations related to career planning**
3	If I graduate with my current major, I will be more likely to get a well paying job
6	If I know my interests and abilities, then I will make better career decisions
7	If I earn my undergraduate degree, I will be more likely to meet my financial goals
15	If I succeed at getting my intended degree, I will be more likely to achieve my career goals
Cluster 3	**Outcome expectations related to success in lab**
9	If I try and understand the chemistry while performing an experiment, I will do well in laboratory
20	If I finish my experiment and while in lab, figure out what my data means, I expect to do well in laboratory
23	If I understand the principles behind the experiments, I will be more likely to succeed in laboratory
Cluster 4	**Outcome expectations related to understanding chemistry**
18	If I understand a fundamental concept, I can solve homework/exam problems on that concept
21	If I can explain a problem or concept to a classmate, I will understand the material better
Cluster 5	**Outcome expectations related to learner based tasks**
4 •	If all I do is memorize the solution to any problem, I will be successful in this course
11 •	If I don't understand the concepts in this course, I can pass (with at least a C)
25 •	If I can remember the solution to a problem and know where to put numbers, do well on quizzes/exams

Despite being a feeder and terminal course respectively, the formation of meaningful clusters from highly heterogeneous datasets in these two courses indicates that the COES could be a viable survey to measure outcome expectations in courses similar to general chemistry.

Limitations

Although self-report surveys to measure affective constructs have been used extensively, any self-report instrument cannot guarantee the absence of response bias. While the COES items did not demonstrate ceiling or floor effects, the COES also attempted to reduce acquiescence bias by including negatively worded tasks.

One of the concerns regarding existing measures of outcome expectations has been the lack of both positive and negative potential outcomes (Fouad & Guillen, 2006). Due to the positively worded tasks in the CSEAS, the COES includes negative conditions but mostly positive outcomes, thus sustaining a limitation mentioned for current outcome expectations surveys.

Despite the use of interviews to offer validity for some of the quantitative results, interpretations and reasons provided in the analyses do not take into account some of the other contextual and environmental factors, such as socioeconomic status and race, that could potentially impact outcome expectations. With the development of an instrument to measure OE, path analyses or structural equation modeling could provide a more thorough exploration of the construct and its relationship to other affective and contextual variables in chemistry.

While this dissertation focused primarily on developing and examining the psychometrics of the COES in first year chemistry courses, utilizing and validating this instrument for use in upper level chemistry courses would perhaps offer more insight into students' expectations about their vocational choices and goals. In addition, testing the validity of the antecedent model using self-efficacy as a precursor will provide a longitudinal view of the relationships among SCCT constructs.

Conclusions and Implications

This study presented a detailed description of the process involved in developing and validating an instrument to measure outcome expectations in chemistry. Exploratory factor analysis of the 25-item instrument resulted in exclusion of seven items and a psychometrically distinct five-factor solution whose fit was tested using confirmatory factor analyses. This model showed a reasonable fit at pre-GCI, post-GCI and pre-GCII time points making this survey sufficiently stable to make longitudinal measurements. Cronbach's alpha for the 18-item scale was .770 while reliability estimates for the subscales ranged from .60 to .81. Validity for the COES was supported in several ways.

The final exam score was significantly correlated to pre-subscales (except OE related to career and success in lab) while performance in the course was significantly correlated to all pre-subscales except OE related to career. Low to moderate correlations were observed suggesting that the COES was not just another measure of academic performance. These correlations support SCCT's hypothesis of a positive relationship between positive OE and academic performance. When assessed using post subscale scores, the correlations became stronger either as a result of OE being measured at a time point close to the final exam or due to the partial influence of self-efficacy beliefs. Gender-based differences were observed for subscales that measured outcomes for tasks very specific to the course (performance- and learner-based tasks, success in lab).

Students who were high performers on the final exam consistently demonstrated more positive outcome expectations for all subscales. Students with more positive expectations related to career outcomes were more certain of persisting in their majors. In addition, students who demonstrated more positive outcome expectations related to performance based tasks and understanding chemistry were more likely to apply material they learned in class to other areas. Career related outcomes did not display significant correlations consistently; being environmental contingencies, their relationship to performance indicators and other variables in the COES was fairly dynamic. It is possible that career related outcomes are impacted by contextual variables that have not been accounted for in these correlations. When CSEAS and COES subscales were correlated, subscales which displayed commonalities in each survey showed significant correlations, indicating support for convergent validity. The ability of the COES to measure fairly distinct dimensions in preparatory chemistry and general chemistry for engineers offers support for the viability of this instrument to measure outcome expectations in courses other than general chemistry – courses which constitute a highly heterogeneous group of students.

Previously developed OE instruments have been used to test and empirically support SCCT hypotheses in science and engineering. While the CHEMX has measured cognitive expectations, the COES is the first instrument to measure outcome expectations in a specific subject such as chemistry. Although the COES attempted to include statements assessing various types of outcome expectations, some of these items – related to self-evaluative tasks or conditions eliciting social approval – did not load on any factor resulting in their exclusion. As no causal relationships were tested, excluding these items does not automatically imply their lack of predictive utility.
This valid and reliable survey offers chemical education researchers a way to capture a fairly unexplored construct, thereby allowing for research into the role of outcome expectations in SCCT, relating this construct to areas of career development theory and subsequently using these results to propose and implement interventions that will help understand students' vocational choices and goals. The meaningful factor structure obtained using the COES is the first step towards making

substantive measurements longitudinally. Psychometric testing is ongoing for this survey and as is the norm for any study that utilizes assessments to capture cognitive or affective measures, using this survey in domains besides chemistry or on a new student demographic may require revisions to survey items and a definite psychometric reevaluation of data resulting from survey administration.

CHAPTER 6: DEVELOPMENT AND VALIDATION OF A SHORTENED INSTRUMENT TO MEASURE SELF-EFFICACY AND OUTCOME EXPECTATIONS

This chapter describes the development and validation of a shortened instrument to measure finer changes in self-efficacy and outcome expectations when administered at key points throughout a single semester.

Background and Rationale

Shortened instruments have been employed to capture affective dimensions in chemistry in an effort to either curb student fatigue that possibly accompanies the usage of long surveys or to examine how affective measures vary on a narrower and frequent time scale (several time points across a semester) as opposed to just the start vs. end time points. Xu and Lewis (2011) utilized factor analysis to refine and shorten the ASCI (Bauer, 2008). The original 20-item ASCI was administered to a different group of students and resulting data were analyzed using EFA and CFA in an effort to replicate the original results. Following this process, the researchers conceptualized new scales based on psychometric evidence and conducted CFA on the newly reconstructed models. According to Xu and Lewis (2011), items with poor descriptive measures, especially skew and kurtosis, low item-total correlations, weak factor loadings or strong factor loadings that cross load elicit removal from an instrument. By using CFA to test several one- and two-factor combinations of items from the original factors, an 8-item instrument – ASCI(V2) – was developed and validated.

Using the shortened ASCI version, in conjunction with surveys to measure self-concept and motivation, Chan and Bauer (2014) employed cluster analyses to identify at-risk students in general chemistry. Six affective variables - measuring self-concept, self-efficacy, anxiety and other attitudinal dimensions – from three surveys were used to categorize students into low, medium and high groups based on their scores on each variable. The predictive utility of these variables was supported by differences being observed between high, medium and low scoring groups on all variables. Students in the high cluster group demonstrated better study strategies and displayed significantly higher performance on the first hourly exam than medium and low cluster groups.

Using five items – assessing self-efficacy beliefs regarding applying chemistry knowledge – from the CAEQ, Villafãne et al., (2014) measured chemistry self-efficacy five times during a semester in a preparatory chemistry course for science majors. Based on CFA results, the self-efficacy items were interpreted as measuring one construct; consequently, analyses were conducted at the scale level (using a composite self-efficacy score) as opposed to at the item level (Villafãne, 2014). With the aid of multilevel modeling (MLM) changes in self-efficacy were item-level self-efficacy were examined across the semester. In addition, differences based on sex and race/ethnicity were also assessed. Results showed that the apparent differences in expected self-efficacy at the start of the semester were unnoticeable by the end of the semester. More importantly, this study revealed key trends in self-efficacy based on sex and race/ethnicity. These studies emphasize the need for assessment tools that can effectively measure affective constructs over time in order to examine changes overall and by student subgroups.

The results described in chapters 4 and 5 support that the CSEAS and COES capture meaningful data related to self-efficacy and outcome expectations respectively and these measures are not equivalent between instruments or between gender subgroups. Additionally, the significant changes identified over the course (as measured pre to post) of a single semester for factors resulting from the CSEAS and COES indicate the possibility of more changes occurring during the semester (prior to or following performance events). Capturing these changes on a much finer level and identifying the key points at which an affective component drops and results in any corresponding change in persistence is crucial in the development of a comprehensive affective profile for at-risk students; furthermore, disadvantages for female students due to gateway course performance can be investigated through changes in self-efficacy and outcome expectations and this may occur differentially for female students. These persistence profiles will provide opportunities to implement targeted interventions to offset changes in persistence. The first step towards assessing these constructs on a finer level is to employ a shortened instrument that offers simultaneous measurements of self-efficacy and outcome expectations and can be administered at multiple points during a semester. Thus, two objectives guided this study:

a) To construct and validate a subset instrument measuring self-efficacy and outcome expectations.
b) To implement the instrument and resulting data in developing predictive performance models at key performance events during the semester.

Methodology

This section describes the process used in the development of the subset instrument. The selection of items, construction of the instrument, testing and participants will also be detailed. In addition, the analyses conducted to evaluate and validate the resulting data will also be examined.

Item selection for the subset survey

Using the subscales resulting from the full-length surveys – CSEAS and COES – items for the subset were selected based on a) subscales that showed significant changes over a single semester not just overall but also by student subgroups and b) relevance to performance events (testing) or decision points (such as dropping the course) over a semester. In addition to the subscales, item selection for the subset survey was further aided by examining student interviews utilized in the validation of the full-length surveys, descriptive statistics of the items, reliability of the subscales and item-total correlations (ITCs). The final subset instrument comprised of items that were selected using a combination of these factors. **Tables 6.1** and **6.2** show the subscales resulting from the CSEAS and COES respectively.

Table 6.1. Subscales resulting from the CSEAS. Subscales with 'x' showed significant pre-post changes.

Factor	Pre-post change
1. Expectations related to performance based tasks (n=3)	x
2. Expectations related to understanding chemistry (n=5)	x
3. Expectations related to career planning (n=4)	x
4. Expectations related to learner based tasks (n=3)	-
5. Expectations related to success in lab (n=3)	-

Table 6.2. Subscales resulting from the COES. Subscales with 'x' showed significant pre-post changes.

Factor	Pre-post change
1. Self-efficacy related to assessment and evaluation (n=7)	x
2. Self-efficacy related to interpersonal tasks (n=3)	x
3. Self-efficacy related to problem solving strategies (n=3)	x
4. Self-efficacy related to higher order tasks (n=3)	x
5. Self-efficacy related to applying chemistry to everyday tasks (n=3)	x
6. Self-efficacy related to performing low order tasks (n=3)	x

Using the information in **Tables 6.1** and **6.2** as starting points, items were selected for the subset based on some of the criteria described below. Each criterion is detailed using either the COES or CSEAS subscales as examples for subset item selection.

a) Based on the results in chapter 5, the COES subscales that showed significant changes in average pre- vs. post-test scores were OE related to performance based tasks (factor 1), understanding chemistry (factor 2) and career goals (factor 3). Although OE related to learner based tasks and success in lab did not reveal significant changes on a pre vs. post level, there were significant differences observed in these subscales when subgroups of students were examined either based on their gender or performance on the final exam / in the course. Thus, items from these subscales were included in the subset survey.

b) With regards to the CSEAS subscales, although all subscales showed significant changes in average pre- vs. post-test scores, items for the subset were not selected from all subscales. Based on student interviews, items that constituted the factor that measured self-efficacy related to higher order tasks were described as "not being particularly valuable or impactful on student performance in the course". As a result, items from this subscale were excluded from the subset instrument.

c) When ITCs were examined for the CSEAS subscale related to interpersonal tasks, item 21 ('asking questions in lecture') had the lowest ITC; Cronbach's alpha for this subscale increased from .714 to .831 when this item was removed. This item was also described in student interviews as one in which self-efficacy was highly dependent on "whether the lecture environment was conducive enough to allow students to pose questions"; additionally, self-efficacy in asking questions during lecture was "dependent on whether a student was comfortable speaking up in front of their peers as some students exhibit more confidence during face-to-face meetings". Thus, when selecting items from the CSEAS subscale related to interpersonal tasks, this item was excluded from the subset instrument.

d) In a situation where the original subscale comprised of more than three items, for example the CSEAS subscale related to assessment and evaluation or the COES subscale related to understanding chemistry, selections were made based on items that would have the most relevance to performance events or items whose outcomes would specifically target the dimension represented by the subscale respectively. Thus, in the case of self-efficacy related to assessment, the three items that were selected for the subset were confidence related to preparing for exams, taking exams and receiving a grade on the exam. Similarly, for outcome expectations related to understanding chemistry, the two items selected for the subset were those whose outcomes involved understanding the workings of the physical world and relating chemistry to situations in everyday life.

Given that the original full-length surveys had 3-6 items per subscale (three items being the minimum for each subscale to ensure over-identification of the construct in CFA), at least two items (fairly correlated with each other) from each subscale in the COES and CSEAS were incorporated into the subset instrument to reliably and substantively represent the original subscales.

Structure of the subset instrument

The 25-item subset survey integrated 13 items from five CSEAS subscales and 12 items from five COES subscales; these 25 items were displayed on one page with the Likert-type response format for each construct preserved from the original surveys. The original subscales from each survey and corresponding items used for the subset are shown in **Tables 6.3** and **6.4**.

Table 6.3. COES subscales and their associated subset items.
Items with asterisks indicate negatively coded items.

Factor & Item	Item	Subset items
Factor 1	**Outcome expectations related to performance based tasks**	
1	If I work hard enough, I will be more likely to pass this course	
22	If I do everything possible (for example, review class notes…), I will do well in this course	x
24	If I do everything possible (for example, review class notes…), be prepared for quizzes/exams in this course.	x
Factor 2	**Outcome expectations related to understanding chemistry**	
14	If I learn chem, I expect to change some of my ideas about how the phys. world works	x
16	If I can relate chem to situations in my everyday life, I expect to learn it better	x
18	If I understand a fundamental concept, I can solve homework/exam problems on that concept	x
17	If I figure out what I did wrong on my exam, I will improve my understanding of course material for the next exam	
19	If I can follow my instructor in lecture, I expect to do better in this course	
Factor 3	**Outcome expectations related to career planning and choices**	
15	If I succeed at getting my intended degree, I will be more likely to achieve my career goals	x
3	If I graduate with my current major, I will be more likely to get a well paying job	
6	If I know my interests and abilities, then I will make better career decisions	x
12	If I obtain a good grade in this course, I will have a better chance of achieving career goals	x
Factor 4	**Outcome expectations related to learner based tasks**	
4*	If all I do is memorize the solution to any problem, I will be successful in this course	x
25*	If I can remember the solution to a problem and know where to put numbers, do well on quizzes/exams	
11*	If I don't understand the concepts in this course, I can pass (with at least a C)	x
Factor 5	**Outcome expectations related to success in lab**	
23	If I understand the principles behind the experiments, I will be more likely to succeed in laboratory	
20	If I finish my experiment and while in lab, figure out what my data means, I expect to do well in laboratory	x
9	If I try and understand the chemistry while performing an experiment, I will do well in laboratory	x

Table 6.4. CSEAS subscales and their associated subset items.

Factor & Item	Item	Subset items
Factor 1	**Self efficacy related to assessment, evaluation**	
20	Doing well on chemistry course exams, given you exert enough effort	
22	Learning material in chemistry courses where considerable math is involved	
26	Preparing for chemistry exams	x
29	Receiving the grade you desire in this course	x
23	Taking an exam or quiz in your chemistry course	x
24	Taking a chemistry exam or quiz where considerable math is involved	
25	Signing up for more chemistry courses in the future (regardless of the outcome of this course or the requirements for your major)	
Factor 2	**Self efficacy related to interpersonal tasks**	
21	Asking questions during lecture	
28	Talking to your chemistry professor	x
27	Understanding your chemistry professor	x
Factor 3	**Self efficacy related to applying problem solving strategies**	
3	Determining appropriate units for a numerical result	
2	Choosing an appropriate equation to solve a chemistry problem	x
1	Understanding what a written chemistry problem is asking you to do	x
Factor 4	**Self efficacy related to higher order tasks**	
18	Writing a summary of the main points of a television documentary that deals with some aspect of chemistry	
16	Explaining why addition of salt melts ice	
17	Using chemistry to propose a solution that keeps cooking water from boiling over	
Factor 5	**Self efficacy related to applying chemistry to everyday tasks**	
11	Converting your speedometer reading from mph to yards/second (1 mile = 1760 yards)	x
12	Calculating the density of lemonade (made by adding 50g of lemons to 500mL of water)	
13	Identifying the type of change (physical vs. chemical) when milk gets sour	x
Factor 5	**Self efficacy related to performing low order tasks**	
5	Describing trends in the periodic table (atomic size, electronegativity)	x
4	Reading and writing a chemical formula	x
8	Identifying elements that are gases at room temperature (from the periodic table)	
** 14	Calculating the percent composition of iron in rust (Fe_2O_3) obtained from your garage door	x

The two asterisks for item 14 in **Table 6.4** indicate that although item 14 was not part of the final CSEAS model, it was included in the subset survey, at the time, based on student interviews, ITCs and its sporadic loading in the SE factor related to applying chemistry to everyday tasks. However, in order to preserve the integrity of the final CSEAS model, this item was included in factor analyses of the subset instrument to track its presence in factor structures at each point, but was omitted when calculating average subscale scores for further analyses. Thus, only the two items from the SE factor related to applying chemistry to everyday tasks (as indicated in **Table 6.4**) were used as representative items to calculate a composite score for this subscale in the subset instrument. Item 14 was excluded from this calculation.

Student interviews

The solicitation and interview processes were similar to those described in chapter 4 for the CSEAS except students were not given notecards with item names and numbers for the process of creating item groups. 15 students in total signed up to participate in semi-structured think aloud interviews, during which the subset survey was presented and students responded to each statement while verbalizing their thought process. As there was no change in the items themselves, these interviews were conducted to primarily evaluate the design of the survey (presence of items from both constructs on the same page of the survey as opposed to being administered full length surveys at different times) and examine the effectiveness of the prompts as these surveys were administered across the semester. Consequently, students were solicited for interviews twice – before the first and third hourly exams. The first set of interviews were conducted before exam 1 in an attempt to offer fairly "untainted" opinions about the items and survey structure in general. The second set of interviews – conducted before exam 3, were examined for effects of survey familiarity and the context of students' responses, especially for the items related to outcome expectations. Students were primarily biology and microbiology majors, with some of them on a pre-med track. Compensation was a $20 gift card to the university book store; the interviews lasted 45 – 60 minutes.

During the second set of interviews, several students stated that they were retroactively making causal associations between the task and outcome when responding to the outcome expectations items. They also mentioned that the retroactive associations were more distinct for surveys administered closer to the end of the semester. As a result, instead of making changes to the items' outcomes as was done in Chapter 5, a change was made to the prompt for the outcome expectations section of the subset survey administered before and after exam 3. These changes in prompt are shown below:

a) Before and after exams 1 and 2:
   - Please indicate your level of agreement with each of the statements.
b) Before and after exam 3:
   - <u>When thinking about what you will still do in this course</u>, please indicate your level of agreement with each of the statements.

While these changes were made in an effort to guide students into thinking about prospective outcomes, there is a possibility that some students bypassed the prompt entirely and continued responding retroactively.

The final version of the subset instrument is included in **Appendix L**.

Data collection and participants

A schematic indicating the survey administration time points and potential triggers is shown in **Figure 6.1**.

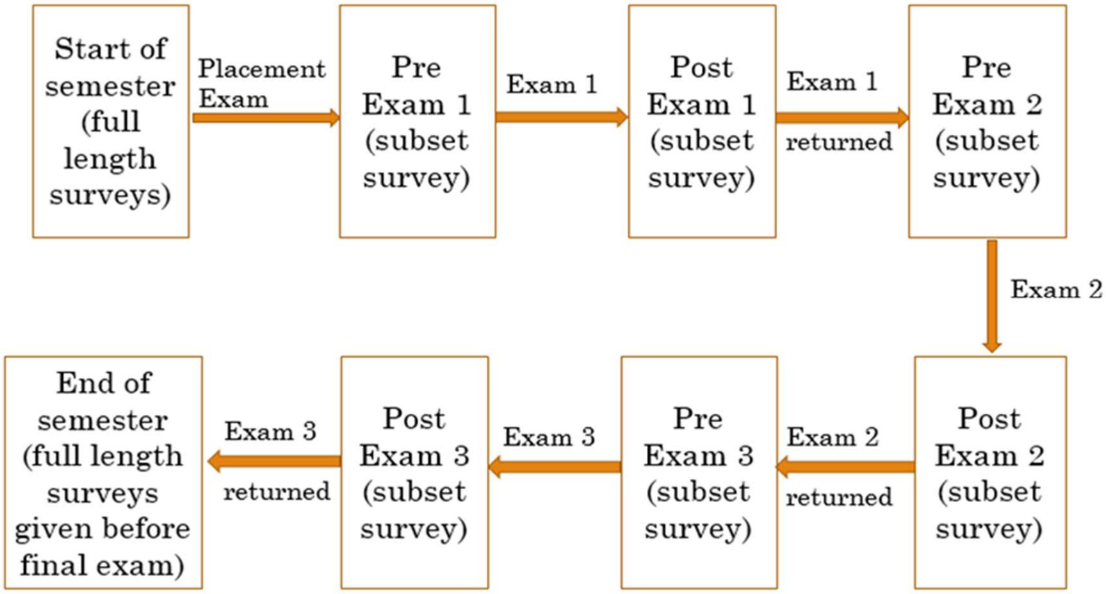


Figure 6.1. Subset administration (GC I and GC II) showing time points and potential triggers

The subset instrument has been administered on paper since Fall 2014; as the goal of the subset survey was to create affective student profiles and offer targeted interventions, data from three semesters (Fall 2014 – Fall 2015) served as the control, pre-intervention dataset.

Subset survey administration took place six times during the semester – before and after exams 1, 2 and 3. The instrument was administered prior to and following these important events throughout the semester to clarify the point at which lower self-efficacy or outcome expectations may occur as well as the events that could trigger this reduction. There was no subset survey administered between exam 3 and the final exam as there were multiple assessments such as the laboratory practical (high stakes) and practice exams (low stakes) that were offered during this interim period to determine the degree to which an event triggered a decrease in self-efficacy or outcome expectations.

The pre- and post-hourly exam surveys were distributed and collected by instructors during the week of the corresponding hourly exam. As an example, for pre-exam surveys, if students had an exam on Wednesday or Thursday, the pre-exam surveys were distributed in lecture on Monday and collected during lecture on Wednesday. Surveys were included in analyses as long as they were returned before students took the exam. Post-exam surveys were distributed and collected during lecture on Friday of the exam week. As grades were posted online or graded exams were being returned during student discussions, it is possible that a subset of students responded to these surveys after they had received their graded exams. While a general review of trends across time points suggest minimal fluctuations in students' affective measures after each exam, the possibility of an interaction between time point and affective measures was not examined in this study.

The surveys typically took 10 minutes to complete and for completing both pre- and post-exam surveys, students were given two extra credit points. The studies described in this chapter were conducted at a large, urban, research intensive public university in the Midwestern United States. To stay consistent with the courses that will ultimately be examined using a longitudinal model, this survey was only administered to the chemistry courses that constituted the two-semester gateway sequence – GC I and GC II. The descriptions of these courses are given in chapter 3. As the interventions that were developed have been tested on GC I students only, the analyses and results presented in this chapter will focus on GC I to help maintain a meaningful transition from a control (pre-intervention) group to the post-intervention group, whose data and results will be the focus of chapter 7. **Table 6.5** shows the number of students – in total and parsed out by gender – who took the surveys at each point. While these numbers are respectable when each time point was compartmentalized, there were only 84 students (28 males and 56 females) who responded to all surveys from start to end.

Table 6.5. GC I participants (by gender) for different time points of subset administration

Fall'14 - Fall'15	Start	Pre Ex 1	Post Ex 1	Pre Ex 2	Post Ex 2	Pre Ex 3	Post Ex 3	End
Total	445	474	448	451	419	417	403	289
Males	208	227	207	207	197	189	187	122
Females	236	247	241	244	222	228	216	167

Data analyses

Data were cleaned as described in chapter 3. Outcome expectations statements that were negatively worded in the full-length surveys and selected for the subset instrument were reverse coded, just as they had been in the full-length surveys. Descriptive statistics were obtained for subscales in the subset survey for assessments of univariate normality, skew, kurtosis and missing data.

Comparative statistics

Similar to the full-length surveys, average subscale scores were calculated for each representative subscale in the shortened, subset instrument. These scores were calculated using raw responses to the items constituting each subscale. In order to facilitate effective comparisons between subscale scores across all time points, including start and end of the semester, the subscales in the full-length surveys had to be "similar" in constitution to the subset instrument. Thus, each average subscale score for the full-length surveys was calculated using just the items that represented the subscale in the shortened instrument as opposed to using all items that comprised that subscale. Interpretation of these scores in the subset survey was similar to the full-length surveys; thus, higher average scores denoted lower self-efficacy and less positive/lower outcome expectations respectively.

For the series of data collected across the semester, comparative statistics involved examining mean subscale scores overall and by gender to elucidate trends in the corresponding affective components across the semester. This was merely a cursory examination of trends, with no implications of significant differences overall, between scores or interactions between gender and subscale scores at each time point.

Psychometric testing – construct validity

Exploratory factor analysis (EFA) was conducted on data obtained at each time point and resulting factor structures were compared to not only evaluate construct validity but also assess concerns regarding over-sampling of students due to repeated administrations of a survey within a short time span. Examining the robustness of the survey constructs was highly important especially in later administrations of the survey.

Predictive validity – Standard multiple linear regression (SMLR)

Given that changes in affective measures (self-efficacy and outcome expectations) could indicate students who are at risk due to gateway course performance, this idea was extended to the subset survey by assessing the impact of affective measures on exam performance. On a very "local" and much finer level, this process entailed developing and testing performance models using affective and cognitive variables and examining their contribution towards predicting student performance on each high-stake assessment (hourly exam) offered during a semester.

Multiple regression is a statistical method used to explore relationships among multiple variables in a sample with the goal of either comprehending a trend and extending this understanding to a population or using a sample to generate a stable regression equation which can be used to predict outcomes for individuals in a different sample (Osborne, 2000). In general, the multiple regression equation of 'Y' on $X_1, X_2, \ldots X_n$ is given by:

$$Y = b_0 + b_1X_1 + b_2X_2 + \ldots + b_nX_n$$

Where b_0 is the intercept and b_1, b_2, b_3, ….b_n are similar to the slope in linear regression equation and are also called regression coefficients.

This study utilized multiple regression for two purposes: To understand how much exam performance was impacted by affective and cognitive variables and to use a combination of these variables (predictors / independent variables) in the generation of a regression equation which was then used to predict students' scores (outcome / dependent variable) on each hourly exam, thus providing evidence of a valid model. Establishing the predictive utility of this model was essential in testing the impact of interventions, whose design and implementation will be described in chapter 7.

In this study, three models were developed to predict performance on three hourly exams. The predictor and outcome variables for each model are summarized in **Table 6.6**.

Table 6.6. Predictors and outcome variable for each subset performance model (GC I)

Model	Predictors	Outcome
1	ACT scores (Composite, Math, Sci-Re) TP Math (Total, Chemistry, Math) Five SE mean subscale scores - Start of semester survey Five OE mean subscale scores - Start of semester survey Five SE mean subscale scores - Pre-Ex 1 subset survey Five OE mean subscale scores - Pre-Ex 1 subset survey Gender	Score on Exam 1
2	Five SE mean subscale scores - Post-Ex 1 subset survey Five OE mean subscale scores - Post-Ex 1 subset survey Five SE mean subscale scores - Pre-Ex 2 subset survey Five OE mean subscale scores - Pre-Ex 2 subset survey Gender Exam 1 score	Score on Exam 2
3	Five SE mean subscale scores - Post-Ex 2 subset survey Five OE mean subscale scores - Post-Ex 2 subset survey Five SE mean subscale scores - Pre-Ex 3 subset survey Five OE mean subscale scores - Pre-Ex 3 subset survey Gender Exam 2 score	Score on Exam 3

The correlation matrix, which displays Pearson's correlation coefficients among the outcome and predictor variables, served as a starting point to decide which predictors (based on smallest p-value or largest t-value) might contribute to the model and account for significant portions of variance. In addition, if the correlation between two predictors was stronger than the correlation between each predictor and the outcome variable, partial correlations were also considered to control for confounding variables. For example, in this case, using model 1 from **Table 6.6** as an example, including an affective subscale from the start of the semester and from the subset survey might not seem particularly beneficial as there could be shared variance between these two predictors that are essentially measuring the same construct as two different time points. Thus, a partial correlation was run for each of the significant predictors, while controlling for co-variates, with the outcome variable to assess the relative impact of each predictor on the outcome (Van den Burg & Lewis, 1988; Soofi et al., 2000).

While regression techniques offer various methods to enter and select predictor variables for the equation (enter, stepwise, forward selection), this study used the standard method of entry and all independent variables were entered into the equation simultaneously. While stepwise regression is recommended when there are several potential predictors in the model, this method was not attempted due its tendency to capitalize on chance and result in biased regression coefficients and variance values (Cook & Weisberg, 1999; Field, 2009).

Some of the assumptions that need to be met for multiple regression include the normal distribution of variables, the existence of a linear relationship between the independent and dependent variable(s), absence of multicollinearity and the normality of residuals. While these are criteria to consider when assessing predictors, the distributional assumption for model errors (residuals) was particularly important as the goal of this study was to use multiple regression to develop and test the fit of a predictive model (Pedhazur, 1997) whose residuals would be examined for changes due to interventions.

Predicting group membership - Discriminant analysis

Discriminant analysis (DA) is a technique used to build a predictive model for group membership. Using a discriminant function, based on a composite (linear combination) of independent (predictor) variables, this model determines the most parsimonious way to achieve maximum separation between two or more naturally occurring and mutually exclusive groups (Klecka, 1980). Discriminant functions are determined using an optimal combination of variables so that the first function maximizes the difference between the values of the outcome (dependent) variable while the second function provides the maximum separation while controlling the first function. Similar to multiple regression, a discriminant score can be calculated based on the weighted combination of the independent variables:

$$\text{Discriminant score} = a + b_1X_1 + b_2X_2 + \ldots + b_nX_n$$

Where b is the discriminant coefficient and X is the independent variable. The coefficients denote the unique contribution of each variable to the discriminant function (Klecka, 1980). These discriminant scores can be divided into each grouping category (low / medium / high or good / bad) and the mean discriminant score can be calculated for each group. The group means on the composite variable are known as centroids. Discriminant coefficients are chosen so as to maximize the distance between centroids and discriminate between the groups to the highest extent (Tabachnick & Fidell, 2001).

Testing the significance of a set of discriminant functions takes place using a matrix of variances and covariances. F tests are used to compare these matrices and determine the existence of significant group differences (with regard to all variables). An overall significant F test leads into further examination of specific variables that have significantly different means across the groups. Once a discriminant function is obtained, discriminant scores can be calculated for new cases which can then be classified into categories in which they had the highest classification scores. DA assumes multivariate normality, homogeneity of variances/covariances, absence of outliers and non-multicollinearity (Klecka, 1980).

In this study, the purpose of DA was to predict student membership in high vs. low performance groups on each hourly exam. As DA for two groups is conceptually similar to multiple regression, both techniques were used in this study as confirmatory methods expected to yield similar results. Despite differences in computations and type of results obtained, regression and discriminant coefficients are interpreted similarly (Kort, 1973). Discriminant functions and regression equations involve linear combinations of the independent variables and their weights, in addition to a constant. While multiple regression was used to predict actual performance on the exam, the use of a DA model provided an index, which - depending on whether its numerical value was above or below a certain point – predicted membership in one of two performance groups. Thus, discriminant analysis was performed in succession to multiple regression analysis, as a secondary confirmation and test of an empirical performance model.

The discriminant function obtained at each performance event in this study was used to classify new cases, exposed to an intervention, in chapter 7. Predictors for each DA model were the same as those used multiple regression; dummy coding was used for the gender variable (0 = male; 1 = female). As no special relationship was predicted or assumed to exist among the predictors and group membership, all predictor variables were entered and analyzed at once, following which only variables which resulted in a significant overall F test were entered. In this study, as the outcome variable was dichotomized, Pearson correlations were interpreted as point-biserial correlation coefficients under the assumptions of approximate normal distributions and lack of outliers for the continuous variable for each category of the dichotomous variable.

Descriptive statistics were obtained for relevant variables at each time point. These analyses were performed using SPSS statistical software 23/24 and Excel 2015/2016.

Results and Discussion

Prior to examining comparative statistics and the variables that impact performance and membership in performance groups, descriptive statistics were obtained for the subscales at each time point. **Tables 6.7** and **6.8** show descriptions, means, standard deviations and other statistics for self-efficacy and outcome expectations subscales respectively. While the sample sizes for OE and SE subscales at each time point are similar as both constructs were part of the same subset survey, the variations in sample size at each point within a construct indicate that not all students completed the series of surveys from start to end.

Table 6.7. Descriptive statistics for self-efficacy subscales at each time point during subset survey administration (F14 – F15) in GC I

Subset factor	Time point	N	Min	Max	Mean	Std. dev.	Skew	Kurtosis
SE - assessment	Start	445	1.00	5.00	2.06	.76	.87	1.06
	Pre Ex 1	474	1.00	5.00	2.58	.92	.61	-.08
	Post Ex 1	448	1.00	5.00	2.71	1.00	.26	-.54
	Pre Ex 2	450	1.00	5.00	2.82	1.00	.21	-.65
	Post Ex 2	418	1.00	5.00	3.09	1.07	.05	-.87
	Pre Ex 3	416	1.00	5.00	2.99	1.01	.15	-.61
	Post Ex 3	403	1.00	5.00	3.03	1.07	.29	-.88
	End	289	1.00	5.00	2.88	1.06	.31	-.83
SE - interpersonal	Start	444	1.00	5.00	1.93	.79	.82	.94
	Pre Ex 1	473	1.00	5.00	2.50	1.05	.56	-.39
	Post Ex 1	447	1.00	5.00	2.51	1.07	.55	-.25
	Pre Ex 2	447	1.00	5.00	2.61	1.06	.34	-.45
	Post Ex 2	416	1.00	5.00	2.71	1.08	.27	-.57
	Pre Ex 3	414	1.00	5.00	2.72	1.04	.36	-.33
	Post Ex 3	401	1.00	5.00	2.76	1.11	.34	-.55
	End	289	1.00	5.00	2.65	1.11	.55	-.36
SE - applying chem. to everyday tasks	Start	445	1.00	5.00	2.01	.79	.73	.37
	Pre Ex 1	474	1.00	5.00	1.84	.72	1.13	1.67
	Post Ex 1	444	1.00	5.00	1.75	.68	.96	1.19
	Pre Ex 2	449	1.00	5.00	1.68	.72	1.36	2.15
	Post Ex 2	417	1.00	5.00	1.74	.71	1.14	1.80
	Pre Ex 3	416	1.00	5.00	1.71	.72	1.09	1.24
	Post Ex 3	400	1.00	4.67	1.73	.73	.96	.65
	End	289	1.00	4.33	1.64	.68	1.25	1.77
SE - low order tasks	Start	443	1.00	4.50	2.03	.70	.67	.38
	Pre Ex 1	474	1.00	5.00	2.11	.80	.75	.35
	Post Ex 1	446	1.00	5.00	2.08	.74	.79	.91
	Pre Ex 2	449	1.00	5.00	2.08	.76	.72	.75
	Post Ex 2	419	1.00	5.00	2.15	.78	.60	.28
	Pre Ex 3	417	1.00	5.00	2.08	.76	.79	.76
	Post Ex 3	402	1.00	4.50	2.05	.76	.73	.29
	End	289	1.00	5.00	1.92	.69	.78	1.07
SE - applying general chem. strategies	Start	445	1.00	4.50	2.23	.77	.68	.13
	Pre Ex 1	474	1.00	5.00	2.08	.75	.97	1.11
	Post Ex 1	448	1.00	5.00	2.05	.75	.82	.91
	Pre Ex 2	450	1.00	5.00	2.11	.78	.70	.53
	Post Ex 2	419	1.00	5.00	2.16	.84	.71	.28
	Pre Ex 3	417	1.00	5.00	2.18	.80	.69	.32
	Post Ex 3	403	1.00	4.50	2.15	.77	.73	.55
	End	289	1.00	4.00	2.00	.72	.66	.27

Table 6.8. Descriptive statistics for outcome expectations subscales at each time point during subset survey administration (F14 – F15) in GC I

Subset factor	Time point	N	Min	Max	Mean	Std. dev.	Skew	Kurtosis
OE - career	Start	445	1.00	3.00	1.55	.49	.55	-.52
	Pre Ex 1	474	1.00	4.33	1.77	.58	.83	1.73
	Post Ex 1	448	1.00	4.00	1.75	.55	.49	.31
	Pre Ex 2	450	1.00	4.00	1.81	.59	.66	.94
	Post Ex 2	417	1.00	4.67	1.83	.63	.78	1.09
	Pre Ex 3	416	1.00	4.00	1.84	.58	.41	.19
	Post Ex 3	402	1.00	4.00	1.82	.62	.50	-.03
	End	289	1.00	4.00	1.78	.55	.48	.45
OE - performance based tasks	Start	445	1.00	5.00	1.38	.56	1.72	4.68
	Pre Ex 1	474	1.00	5.00	1.76	.79	1.23	2.09
	Post Ex 1	448	1.00	5.00	1.82	.82	.92	.68
	Pre Ex 2	450	1.00	5.00	1.94	.87	.80	.36
	Post Ex 2	417	1.00	5.00	2.18	1.02	.78	.08
	Pre Ex 3	416	1.00	5.00	2.11	.87	.82	.92
	Post Ex 3	403	1.00	5.00	2.20	1.07	.78	.01
	End	289	1.00	5.00	1.88	.76	1.02	1.75
OE - learner based tasks	Start	445	1.00	5.00	2.39	.81	.75	.60
	Pre Ex 1	474	1.00	5.00	2.37	.75	.66	.36
	Post Ex 1	448	1.00	5.00	2.40	.75	.59	.24
	Pre Ex 2	451	1.00	5.00	2.39	.77	.60	.31
	Post Ex 2	418	1.00	5.00	2.43	.82	.54	.00
	Pre Ex 3	417	1.00	5.00	2.46	.86	.37	-.32
	Post Ex 3	403	1.00	5.00	2.45	.84	.45	-.09
	End	289	1.00	5.00	2.49	.80	.35	-.15
OE - lab success	Start	445	1.00	4.50	1.59	.55	.82	1.24
	Pre Ex 1	474	1.00	4.50	1.86	.62	.86	1.97
	Post Ex 1	448	1.00	4.00	1.85	.60	.42	.37
	Pre Ex 2	451	1.00	4.00	1.89	.61	.63	1.09
	Post Ex 2	417	1.00	5.00	1.96	.70	.98	2.15
	Pre Ex 3	417	1.00	4.50	1.97	.62	.34	.41
	Post Ex 3	403	1.00	4.50	1.95	.70	.71	1.02
	End	289	1.00	5.00	1.91	.61	.86	3.04
OE - understanding chem.	Start	444	1.00	3.33	1.79	.51	.12	-.64
	Pre Ex 1	474	1.00	4.33	1.98	.59	.32	.50
	Post Ex 1	448	1.00	4.33	1.98	.62	.39	.50
	Pre Ex 2	450	1.00	5.00	2.02	.66	.62	1.30
	Post Ex 2	417	1.00	4.33	2.10	.67	.52	.64
	Pre Ex 3	417	1.00	4.33	2.12	.68	.29	.24
	Post Ex 3	403	1.00	4.33	2.07	.69	.37	.13
	End	289	1.00	4.33	2.04	.56	.40	1.14

Descriptive statistics indicate subscales with high skewness and kurtosis values at certain time points. Given the affective measures and triggers between time points, it is possible that students indicate disproportionately higher or lower perceived confidence and expectations at certain time points relative to others, resulting in skewness and kurtosis values being out of range. While these are criteria to consider when assessing predictors for both analyses, the more important distributional assumption for multiple regression is for model errors; although some subscales were non-normal at each level of the outcome variable for DA, these analyses were carried out under this limitation and without any transformation to the subscales.

When the subscale means for each construct were examined by gender and overall at each time point, no apparent differences were observed between males and females across time points for most subscales. **Figures 6.2** and **6.3** indicate the average self-efficacy and outcome expectations subscale scores at each time point across the semester. As displayed, only 84 complete sets of responses were available to track across the semester. **Figure 6.3**, in particular, shows that OE related to learner based tasks displays the most noticeable differences between males vs. females. On a superficial level, these plots indicate that there might be events that trigger a decline in affective measures at certain time points and these decreases could be manifested differentially based on gender.

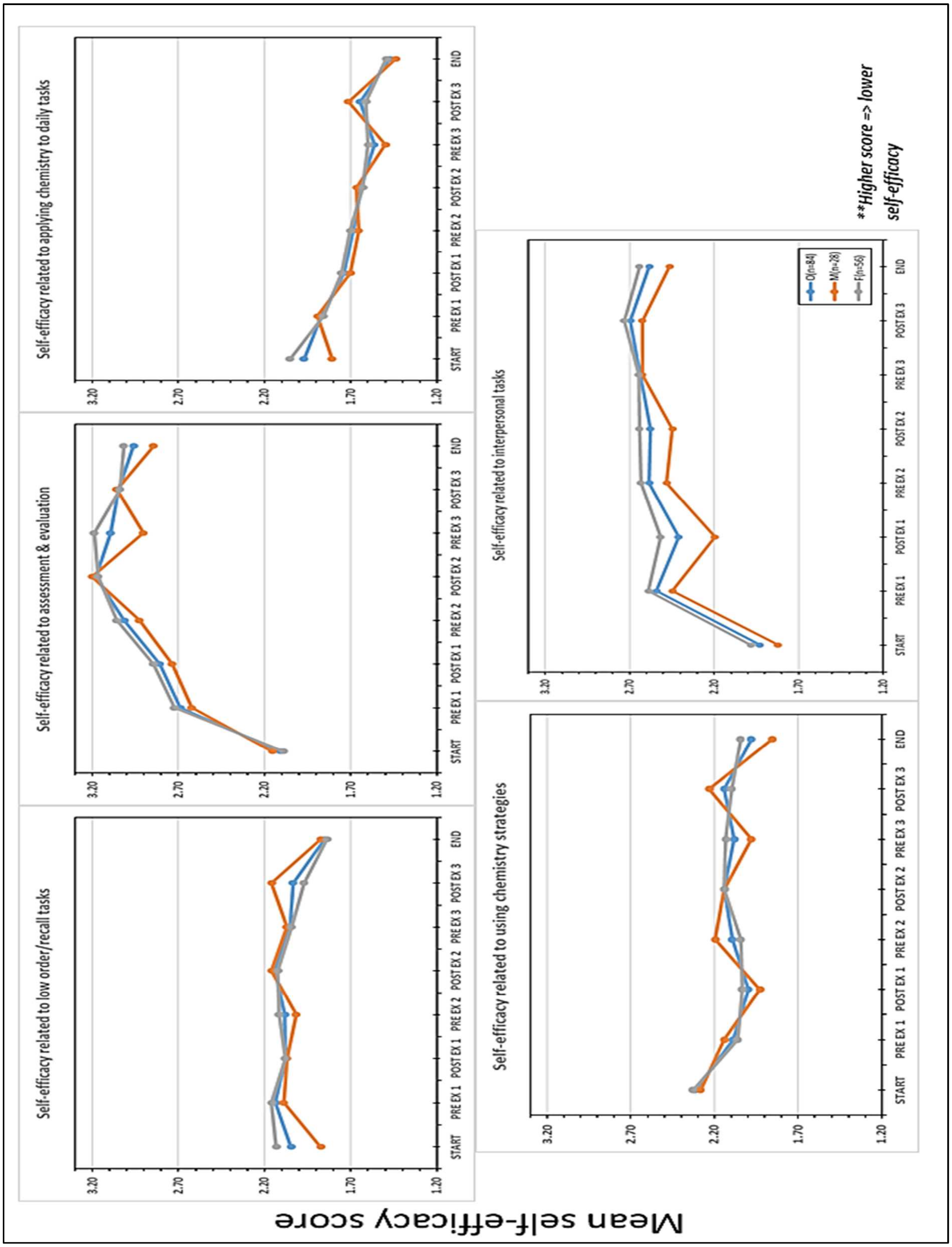


Figure 6.2. Cursory trends in outcome expectations (subset survey) across semesters (AY14-AY15) in GC I

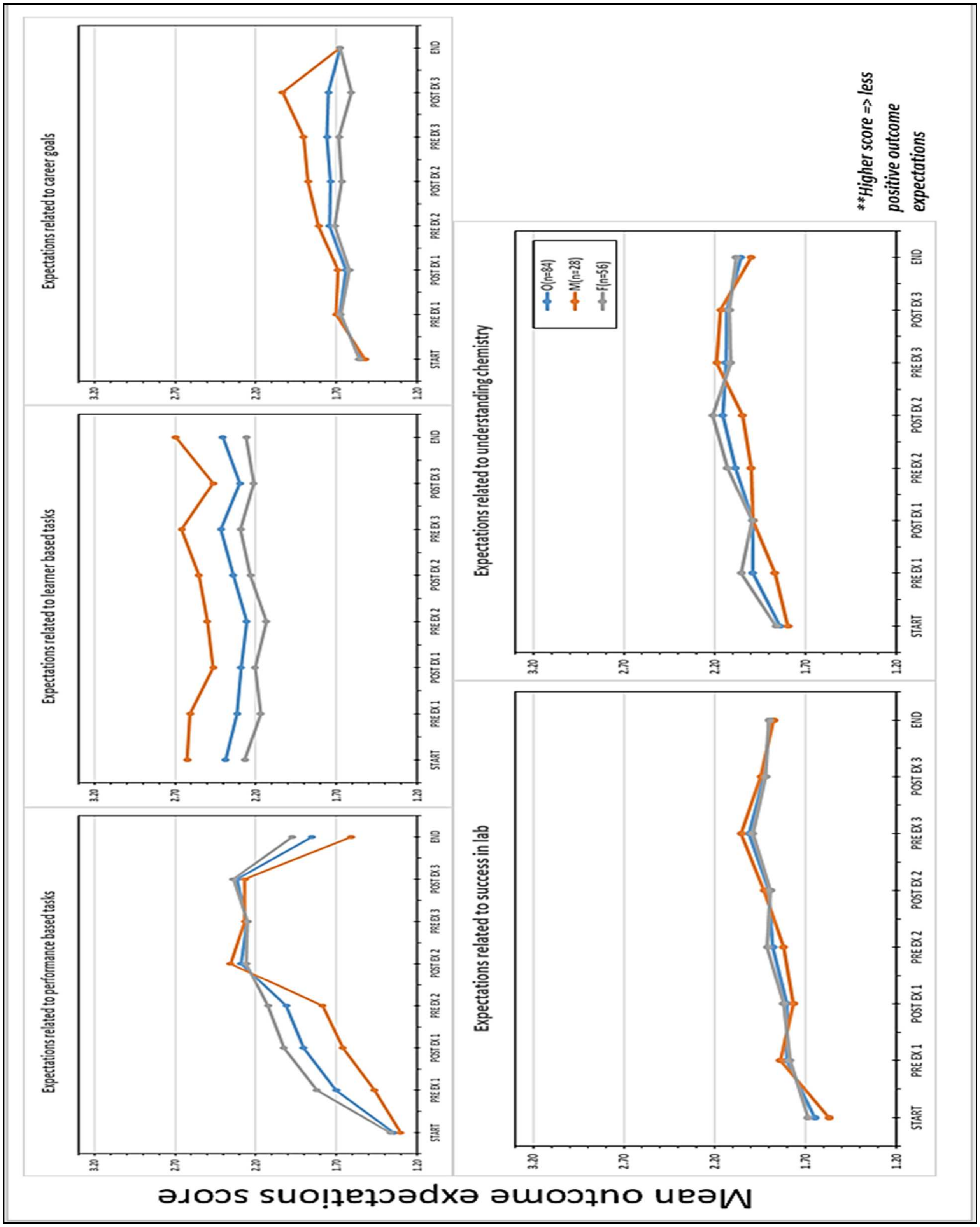


Figure 6.3. Cursory trends in outcome expectations (subset survey) across semesters (AY14-AY15) in GC I

Psychometric testing – Exploratory factor analyses (EFA)

Tables 6.9 and **6.10** show the factor structures for self-efficacy and outcome expectations obtained at each time point across a semester using data from AY14-15. These structures have been displayed by items in sequence (as opposed to size of item loadings in each factor) to facilitate tracking the changes in items in their construct locations across the three time points. Cronbach alpha values for the self-efficacy subscales ranged from .881 - .903 while values for the outcome expectations subscales range from .797 - .834. When comparing the factor structures across the semester and before / after each exam, item movement is observed between factors towards the end of the semester (exam 3) vs. at the start. While certain subscales (items 9-13) stayed grouped consistently, one or two items appeared as standalone items, while others did not load at certain time points. Although interpersonal and assessment items in the self-efficacy instrument were two distinct factors in the full-length surveys, the items selected from these factors grouped together across subset administrations (items 9-13). Subscales in the outcome expectations component of the subset display more stability across a semester, prior to and after performance events. While each structure had some item movement or combination of subscales at different time points, perhaps indicative of oversampling and varied interpretation of items, the presence of distinct, mostly meaningful factors and reasonable reliability estimates suggest a psychometrically valid measure of self-efficacy and outcome expectations.

Table 6.9. Self-efficacy factor structures (tracking item changes) from subset survey administration across a semester (AY14-15) in GC I

Item #	Item in subset	Pre Exam 1 (N = 138) 1	2	3	4	Pre Exam 2 (N=145) 1	2	3	4	Pre Exam 3 (N=146) 1	2	3	4
1	Choosing equation to solve a chem. problem	.517					.707				.753		
2	Determing appropriate units for numerical result		.533				.719				.782		
3	Reading & writing a chemical formula			.669			.705				.695		
4	Describing the trends in periodic table			.854					**.863**				**.900**
5	Identifying type of change when milk gets sour				.929			**.521**	**.507**			.725	
6	Converting speedometer reading		.764					.852				.813	
7	Calculating the density of lemonade		.777					.839				.820	
8	Calculating the % composition of iron in rust		.526										
9	Taking an exam or quiz in your chemistry course	.781				.744				.788			
10	Preparing for chemistry exams	.805				.757				.834			
11	Understanding your chemistry professor	.757				.827				.822			
12	Talking to your chemistry professor	.772				.809				.783			
13	Receiving the grade you desire in this course	.817				.780				.806			

Item #	Item in subset	Post Exam 1 (N=142) 1	2	3	4	Post Exam 2 (N=159) 1	2	3	4	Post Exam 3 (N=124) 1	2	3	4
1	Choosing equation to solve a chem. problem	**.571**	**.558**				.735					.797	
2	Determing appropriate units for numerical result		.842				.788					.614	
3	Reading & writing a chemical formula		.641				.731					.567	
4	Describing the trends in periodic table				.794				**.945**				**.797**
5	Identifying type of change when milk gets sour				.689			.672			.685		
6	Converting speedometer reading							.764			.844		
7	Calculating the density of lemonade			.749				.675			.814		
8	Calculating the % composition of iron in rust			.754				**.541**				**.500**	
9	Taking an exam or quiz in your chemistry course	.847				.751				.779			
10	Preparing for chemistry exams	.777				.758				.832			
11	Understanding your chemistry professor	.763				.756				.797			
12	Talking to your chemistry professor	.751				.817				.847			
13	Receiving the grade you desire in this course	.802				.721				.842			

Table 6.10. Outcome expectations factor structures (tracking item changes) from subset survey administration across a semester (AY14-15) in GC I.

Item #	Item in subset (N = 310)	Pre Exam 1: 1	2	3	4
14	All I do is memorize solution, will be successful in this course				.791
15	Know interests and abilities, make better career decisions	.733			
16	Try and understand chemistry while performing an expt, can pass with C	.745			
17	Don't understand concepts, can pass (with at least a C)				.818
18	Obtain good grade in course, better chance of achieving career goals	.589			
19	Learn chemistry, change ideas about how physical world works			.865	
20	If I succeed at getting degree, more likely to achieve career goals	.740			
21	Relate chemistry to everyday life, expect to learn it better			.721	
22	Fundamental concept, solve any HW/exam prob. On that concept	.635			
23	Finish expt and while in lab, figure out data, expect to do well in lab	.577			
24	Do everything possible, do well in this course		.895		
25	Do everything possible, be prepared for exams / quizzes in this course		.899		

Item #	Item in subset (N = 293)	Pre Exam 2: 1	2	3
14	All I do is memorize solution, will be successful in this course			.808
15	Know interests and abilities, make better career decisions	.675		
16	Try and understand chemistry while performing an expt, can pass with C			
17	Don't understand concepts, can pass (with at least a C)			.788
18	Obtain good grade in course, better chance of achieving career goals	.702		
19	Learn chemistry, change ideas about how physical world works	.699		
20	If I succeed at getting degree, more likely to achieve career goals	.761		
21	Relate chemistry to everyday life, expect to learn it better	.715		
22	Fundamental concept, solve any HW/exam prob. On that concept		.549	
23	Finish expt and while in lab, figure out data, expect to do well in lab		.703	
24	Do everything possible, do well in this course		.884	
25	Do everything possible, be prepared for exams / quizzes in this course		.883	

Item #	Item in subset (N = 282)	Pre Exam 3: 1	2	3	4
14	All I do is memorize solution, will be successful in this course				.773
15	Know interests and abilities, make better career decisions		.836		
16	Try and understand chemistry while performing an expt, can pass with C		.637		
17	Don't understand concepts, can pass (with at least a C)				.844
18	Obtain good grade in course, better chance of achieving career goals		.576		
19	Learn chemistry, change ideas about how physical world works			.853	
20	If I succeed at getting degree, more likely to achieve career goals		.709		
21	Relate chemistry to everyday life, expect to learn it better			.788	
22	Fundamental concept, solve any HW/exam prob. On that concept	.644			
23	Finish expt and while in lab, figure out data, expect to do well in lab	.804			
24	Do everything possible, do well in this course	.884			
25	Do everything possible, be prepared for exams / quizzes in this course	.768			

Item #	Item in subset (N = 328)	Post Exam 1: 1	2	3	4
14	All I do is memorize solution, will be successful in this course				.798
15	Know interests and abilities, make better career decisions		.521		
16	Try and understand chemistry while performing an expt, can pass with C	.533			
17	Don't understand concepts, can pass (with at least a C)				.769
18	Obtain good grade in course, better chance of achieving career goals			.844	
19	Learn chemistry, change ideas about how physical world works		.769		
20	If I succeed at getting degree, more likely to achieve career goals			.850	
21	Relate chemistry to everyday life, expect to learn it better		.826		
22	Fundamental concept, solve any HW/exam prob. On that concept	.656			
23	Finish expt and while in lab, figure out data, expect to do well in lab	.689			
24	Do everything possible, do well in this course	.899			
25	Do everything possible, be prepared for exams / quizzes in this course	.906			

Item #	Item in subset (N = 286)	Post Exam 2: 1	2	3	4	5
14	All I do is memorize solution, will be successful in this course					.864
15	Know interests and abilities, make better career decisions		**.593**	.601		
16	Try and understand chemistry while performing an expt, can pass with C			.822		
17	Don't understand concepts, can pass (with at least a C)					.707
18	Obtain good grade in course, better chance of achieving career goals		.806			
19	Learn chemistry, change ideas about how physical world works				.767	
20	If I succeed at getting degree, more likely to achieve career goals		.829			
21	Relate chemistry to everyday life, expect to learn it better				.746	
22	Fundamental concept, solve any HW/exam prob. On that concept	.595				
23	Finish expt and while in lab, figure out data, expect to do well in lab			.646		
24	Do everything possible, do well in this course	.901				
25	Do everything possible, be prepared for exams / quizzes in this course	.909				

Item #	Item in subset (N = 293)	Post Exam 3: 1	2	3	4
14	All I do is memorize solution, will be successful in this course				.766
15	Know interests and abilities, make better career decisions	.851			
16	Try and understand chemistry while performing an expt, can pass with C	.566			
17	Don't understand concepts, can pass (with at least a C)				.816
18	Obtain good grade in course, better chance of achieving career goals	.742			
19	Learn chemistry, change ideas about how physical world works			.836	
20	If I succeed at getting degree, more likely to achieve career goals	.810			
21	Relate chemistry to everyday life, expect to learn it better			.829	
22	Fundamental concept, solve any HW/exam prob. On that concept		.647		
23	Finish expt and while in lab, figure out data, expect to do well in lab	**.549**	.596		
24	Do everything possible, do well in this course		.912		
25	Do everything possible, be prepared for exams / quizzes in this course		.908		

Standard multiple linear regression (SMLR)

Three regression models were developed to predict the students' performance on three performance events (hourly exams 1, 2 and 3). To achieve a respectable sample size for each model, data were compartmentalized and examined using predictors leading up to each performance event. Due to the reduced number of complete surveys, data were not examined as a time series; instead each model was developed using students who had complete data for predictor and outcome variables utilized for that model. The analyses and results for development of model 1 (for exam 1) have been described in detail. As similar protocols were followed for subsequent performance models, only the results have been included for these models.

<u>Model 1: Predicting performance on Exam 1</u>

One of the first steps to determine the variables that would be included in the model was to conduct a correlational analysis. If a predictor displayed a significant correlation with the outcome variable, it was considered for potential inclusion. The correlation matrix (for this model) displaying significant correlations among variables and the score on Exam 1 is shown in **Table 6.11.**

Table 6.11. Bivariate correlations between predictor variables and score on Exam 1 for GCI (AY14-15). Higher SE/OE scores indicate lower SE/OE.

		ACT_Comp	ACT_Math	ACT_Science	SE_assessment - START	SE_interpersonal - START	SE_everday tasks - START	SE_low order tasks - START	SE_applying strategies - START	OE_Career - START	OE_performance based tasks - START	OE_learner based tasks - START	OE_lab success - START	OE_understanding chem - START	TP - Math	TP - Chem	TP - Total	SE_assessment - Pre Ex 1	SE_interpersonal - Pre Ex 1	SE_everyday tasks - Pre Ex 1	SE_low order tasks - Pre Ex 1	SE_applying strategies - Pre Ex 1	OE_career - Pre Ex 1	OE_performance based tasks - Pre Ex 1	OE_learner based tasks - Pre Ex 1	OE_lab success - Pre Ex 1	OE_understanding chem - Pre Ex 1	Exam 1 score
ACT_Comp	Pearson Correlatio	1																										
ACT_Math	Pearson Correlatio	.792**	1																									
ACT_Science	Pearson Correlatio	.825**	.643**	1																								
SE_assessment - START	Pearson Correlatio	-0.102	-0.086	-0.073	1																							
SE_interpersonal - START	Pearson Correlatio	-0.084	-0.071	-0.098	.635**	1																						
SE_everday tasks - START	Pearson Correlatio	-.138*	-.158*	-0.097	.247**	.224**	1																					
SE_low order tasks - START	Pearson Correlatio	0.083	0.085	0.066	.303**	.193**	.361**	1																				
SE_applying strategies - START	Pearson Correlatio	-.152*	-0.098	-0.089	.410**	.258**	.427**	.430**	1																			
OE_Career - START	Pearson Correlatio	-0.049	-0.019	-0.032	.315**	.243**	0.077	0.038	.165**	1																		
OE_performance based tasks -	Pearson Correlatio	-.142*	-0.104	-.167**	.366**	.262**	.171**	.149*	.193**	.484**	1																	
OE_learner based tasks - START	Pearson Correlatio	-.163**	-.140*	-0.096	-0.093	-0.013	-0.103	-.187**	-0.072	-0.036	0.001	1																
OE_lab success - START	Pearson Correlatio	-0.118	-0.117	-.134*	.313**	.179**	0.110	.130*	0.101	.510**	.567**	-0.033	1															
OE_understanding chem - START	Pearson Correlatio	-0.113	-0.092	-.129*	.371**	.214**	.163**	.140*	.199**	.446**	.376**	-0.054	.467**	1														
TP - Math	Pearson Correlatio	.338**	.412**	.303**	-0.064	-0.040	-.154*	0.115	-0.102	0.022	-0.077	-0.003	-0.073	-0.043	1													
TP - Chem	Pearson Correlatio	.248**	.286**	.252**	-0.015	0.057	-0.110	-0.054	-.138*	-0.067	-0.020	0.023	-0.032	-0.036	.323**	1												
TP - Total	Pearson Correlatio	.370**	.421**	.334**	-0.085	-0.003	-.270**	-0.040	-.194**	-0.014	-0.068	0.021	-0.084	-0.037	.718**	.697**	1											
SE_assessment - Pre Ex 1	Pearson Correlatio	-.168**	-.212**	-.196**	.545**	.262**	.251**	.240**	.264**	0.118	.255**	-.163**	.139*	.135*	-.254**	-0.071	-.268**	1										
SE_interpersonal - Pre Ex 1	Pearson Correlatio	-0.082	-.167**	-0.093	.381**	.371**	.227**	.288**	.246**	0.052	.166**	-0.044	0.095	.138*	-.127*	0.077	-0.112	.645**	1									
SE_everyday tasks - Pre Ex 1	Pearson Correlatio	-.224**	-.219**	-.151*	.309**	.196**	.396**	.316**	.365**	.125*	.232**	0.064	.135*	.153*	-.210**	-.168**	-.276**	.416**	.329**	1								
SE_low order tasks - Pre Ex 1	Pearson Correlatio	0.020	0.021	0.011	.189**	0.010	.164**	.416**	.241**	0.072	0.101	-0.012	0.050	.175**	-0.018	-.147*	-0.117	.378**	.369**	.441**	1							
SE_applying strategies - Pre Ex	Pearson Correlatio	-.148*	-.216**	-0.118	.383**	.253**	.236**	.268**	.384**	.147*	.189**	0.069	0.057	0.081	-.246**	-0.096	-.218**	.544**	.436**	.537**	.497**	1						
OE_career - Pre Ex 1	Pearson Correlatio	-.174**	-.197**	-.150*	.228**	.184**	0.101	0.036	0.116	.446**	.286**	0.107	.278**	.277**	-.140*	-.135*	-.137*	0.048	0.021	.129*	0.031	0.108	1					
OE_performance based tasks - Pre	Pearson Correlatio	-0.092	-0.101	-0.114	.332**	.170**	.162**	.128*	.155*	.228**	.298**	-0.024	.312**	.298**	-.235**	-0.071	-.229**	.283**	.237**	.177**	.141*	.185**	.465**	1				
OE_learner based tasks - Pre Ex 1	Pearson Correlatio	-0.110	-0.049	-0.092	-0.036	0.044	-0.041	-0.055	0.032	0.078	0.054	.437**	0.090	0.058	-0.006	-0.106	-0.019	-.226**	-0.091	0.005	-0.078	-0.019	.149*	0.017	1			
OE_lab success - Pre Ex 1	Pearson Correlatio	-.175**	-.160*	-.182**	.276**	.235**	0.081	0.044	.126*	.357**	.363**	0.055	.407**	.338**	-.162**	-0.048	-.135*	.141*	.123*	0.117	0.055	.138*	.600**	.592**	.124*	1		
OE_understanding chem - Pre Ex 1	Pearson Correlatio	-0.055	-0.095	-0.065	.259**	.163**	.170**	.126*	0.113	.338**	.253**	-0.056	.352**	.485**	-.128*	-.140*	-.184**	0.101	0.047	.123*	.154*	0.113	.549**	.534**	0.004	.621**	1	
Exam 1 score	Pearson Correlatio	.380**	.358**	.344**	-.275**	-.168**	-.252**	-.135*	-.155*	-0.043	-.167**	-0.034	-0.113	-0.093	.422**	.137*	.543**	-.355**	-.242**	-.238**	-0.074	-.248**	-0.096	-.239**	0.047	-0.123	-0.093	1

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).
Listwise N=254

Among the performance indicators, the strongest and significant correlations to score on Exam 1 resulted from the placement measures: ACT Composite ($r=.380^{**}$), ACT Math ($r=.358^{**}$), ACT Sci-Re ($r=.344^{**}$), TP Math ($r=.422^{**}$), TP Chem. ($r=.137^{*}$) and TP total ($r=.543^{**}$). Among the persistence measures, there were significant correlations among almost all SE subscales (at the start and pre-Ex 1) with the highest correlation for SE related to assessment at pre-Ex 1 ($r=-.355^{**}$). Fewer OE subscales at either time point correlated significantly with the score on Exam 1.

Closer examination of the matrix indicated some strong, significant inter-correlations among performance variables; as these inter-correlations were higher than those displayed between these predictors and the outcome variable, and the violation of the assumption of multicollinearity was an issue when considering these predictors. Consequently, only those predictors that displayed significantly stronger correlations with the dependent variable than other variables were considered in the model.

ACT Math and TP total had the strongest correlations with the Exam 1 score ($r=.358^{**}$ and $r=.543^{**}$ respectively); however, ACT Math displayed a higher correlation with TP total ($r=.421^{**}$). Among the affective predictors, SE related to everyday tasks, applying strategies and assessment had moderate correlations with the outcome variable (at the start and pre-Ex 1). However, these variables also had strong inter-correlations as they were measuring the same subscale at two different points. Thus, partial correlations were examined to assess the relative impact of each predictor. For instance, when SE related to assessment at the start was correlated with Exam 1 score, controlling for SE related to assessment at pre-Ex 1, the correlation was no longer significant. However, when SE related to assessment at pre-Ex 1 was correlated with the outcome variable, controlling for the effects of assessment related SE at the start, this correlation was still significant. Thus, using the correlation matrix, in combination with partial correlations resulted in a model with two predictors (total TP scores and SE related to assessment at pre-exam 1) that accounted for 34% of the variance in the model as shown in **Table 6.12.**

Table 6.12. Summary of multiple regression analysis for students' Exam 1 scores in GC I (N=254)

Variables	B	SE (β)	β	t	Sig. (p)	Zero-order	Partial	Part	Tolerance	VIF
TP_total	.761	.084	.483	9.086	.000	.543	.497	.465	.928	1.077
SE_assessment - Pre Ex 1	-3.573	.843	-.225	-4.237	.000	-.355	-.258	-.217	.928	1.077
R^2 = .342; Adj. R^2 = .337 F(2,253) = 65.250, p < .001 **Average of residuals (control) = -0.01**	Exam 1 score = 30.034 + (.761 TP_total) + (-3.573 SE_assessment - Pre Ex 1)									

The model was statistically significant, $F(2,253) = 65.250$, $p < .001$. The unstandardized coefficients (B) provide information about the relationship between the score on Exam 1 and each predictor. In this model, as total TP scores increase by one point, the score on Exam 1 increases by .761 points. Because in the self-efficacy scale, a higher mean subscale score indicates lower self-efficacy, an increase in self-efficacy score related to assessment and exam preparation right before exam 1 by 1 point decreases the score on exam 1 by 3.57 points. As this affective measure was closely tied to the outcome (performance on exam 1), it is expected that it would make a strong contribution to the performance model, in addition to past performance indicators (ability), which students draw upon considerably in learning course material for a performance event.

Placing emphasis on model errors (residuals), standardized regression residual plots, as displayed in **Figure 6.4**, showed most of the residual values around zero with no obvious 'funneling'; thus, homoscedasticity was assumed. Additionally, the average of residuals was zero, normal P-P plot of regression standardized residual did not show deviations from the straight line and Shapiro-Wilk's test of normality was not significant (p=.056), indicating that residuals were normal. Although sum and mean of residuals in a least-squares regression are exactly zero as long as an intercept term is included as this is a consequence of the "normal equations" that are solved to find the estimates of the regression coefficients (Pedhazur, 1997), in this model the average of the residuals was close to zero but the sum was not zero. This could perhaps indicate error in model specification, thus necessitating a change in the model or a closer investigation into the predictors used for model development.

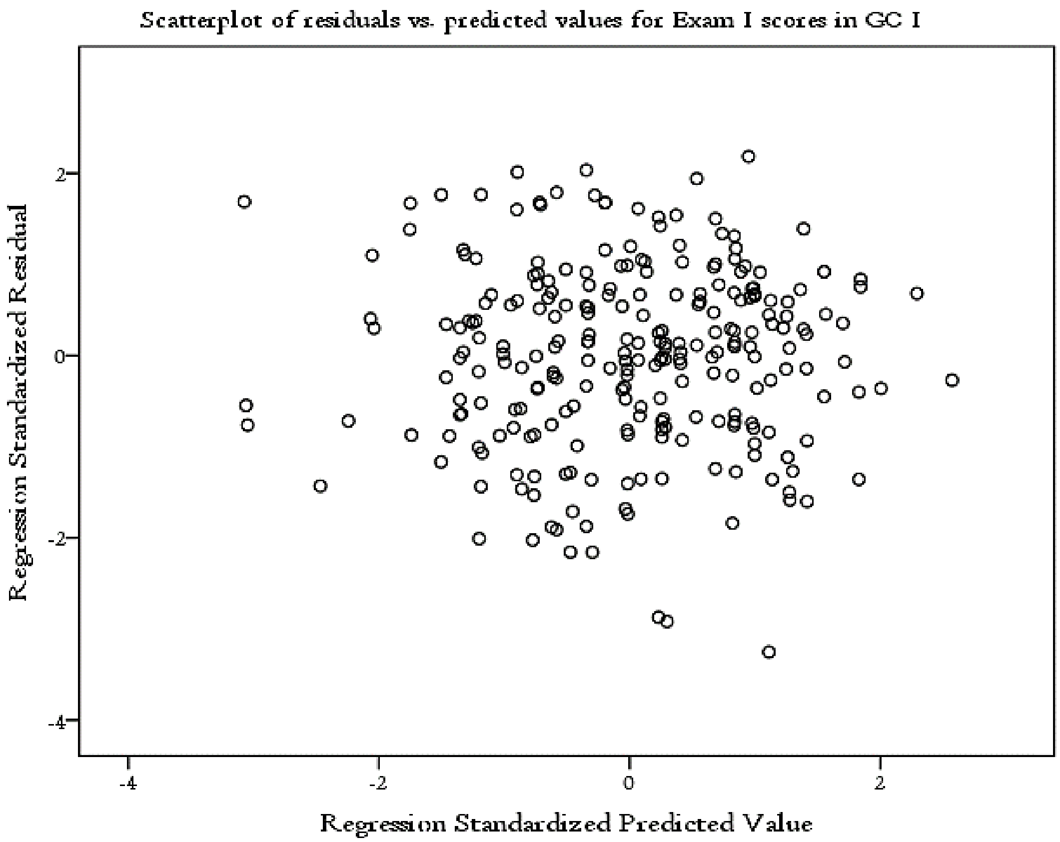


Figure 6.4. Scatter plot distribution of residuals for a performance model using scores on Exam 1 – GCI

Similar models, using the protocols described above, were developed for predicting performance on the second (Exam 2) and third (Exam 3) hourly exams respectively. Only the summary of regression analyses and results of Shapiro-Wilks' normality tests have been included for these models. Scatterplots for each model did not show 'funneling'; in addition, P-P plots also did not display deviations from the straight line.

Model 2: Predicting performance on Exam 2

This model was statistically significant, $F(3,309) = 79.319$, $p<.001$. Three predictors accounted for 43.2 % of the variance in the model. While the score on exam 1 was a significant predictor in this model, affective variables also made significant contributions towards predicting scores on the second exam as shown in **Table 6.13**.

Table 6.13. Summary of multiple regression analysis for students' Exam 2 scores in GC I (N=310)

Variables	B	SE (β)	β	t	Sig. (*p*)	Zero-order	Partial	Part	Tolerance	VIF
Exam 1 score	.630	.051	.575	12.303	.000	.632	.575	.528	.840	1.190
OE_learner based tasks - Post Ex 1	-1.903	.905	-.091	-2.103	.036	-.019	-.119	-.090	.980	1.020
SE_applying strategies - Pre Ex 2	-3.871	.993	-.181	-3.897	.000	-.399	-.217	-.167	.854	1.170

$R^2 = .437$; Adj. $R^2 = .432$ Exam 2 score = 33.109 + (.630 Exam 1 score) + (-1.903 OE_learner based tasks - Post Ex 1)
$F(3,309) = 79.319$, $p < .001$ + (-3.871 SE_applying strategies - Pre Ex 2)
Average of residuals (control) = -0.02

Average of residuals was -0.02; a non-significant result for Shapiro-Wilk normality test on standardized residuals (p=.750) indicated that the residuals were normal. As score on exam 1 increases by one point, score on exam 2 increases by 0.63 points; as post-exam 1 expectations related to learner based tasks increases by 1 point, the score on exam 2 decreases by 1.9 points and as confidence related to applying strategies increases by 1 point, the score on exam 2 decreases by 3.87 points.

Model 3: Predicting performance on Exam 3

The summary of these regression results is displayed in **Table 6.14**. This model was statistically significant, $F(2,297) = 120.020$, $p<.001$. Average of residuals was 0.01; Shapiro-Wilk normality test on standardized residuals was non-significant (p=.703), indicating that residuals were normal. Two predictors accounted for 44.5% of the variance in this model. Increasing the score on exam 2 by a point increases the corresponding score on exam 3 by .61 points, while an increase in the score of confidence related to interpersonal tasks (post Ex-2) by 1 point reduced scores on exam 3 by 1.66 points.

Table 6.14. Summary of multiple regression analysis for students' Exam 3 scores in GC I (N=298)

Variables	B	SE (β)	β	t	Sig. (*p*)	Zero-order	Partial	Part	Tolerance	VIF
Exam 2 score	.608	.044	.627	13.862	.000	.661	.628	.599	.912	1.096
SE_interpersonal - Post Ex 2	-1.659	.662	-.113	-2.507	.013	-.299	-.144	-.108	.912	1.096

R^2 = .449; Adj. R^2 = .445 Exam 3 score = 33.960 + (.608 Exam 2 score) + (-1.659 SE_interpersonal - Post Ex 2)
F(2,297) = 120.020, p < .001
Average of residuals (control) = 0.01

Discriminant analyses (DA)

Data analyses using this method was approached in the same way as multiple regression; due to the reduced number of complete surveys across a semester, each performance event was evaluated as a compartmentalized model in order to utilize a larger dataset leading up to each event. Three DA models were developed to predict student membership in high vs. low performing groups on three performance events (hourly exams 1, 2 and 3). High vs. low performing groups were designated by calculating z-scores for each student's exam score. Z-scores > 0 were denoted as the high performing group while z-scores < 0 were considered the low performing group. The analyses and results for predicting group membership for exam 1 (model 1) have been described in detail. As similar protocols were followed for subsequent models, only the key results have been included for these models.

<u>Model 1: Predicting membership in high vs. low performing groups on Exam 1</u>

Similar to multiple regression, the starting point of this analysis was to find potential predictors for the membership model corresponding to each performance event. As the outcome variable in this case was dichotomous, point-biserial correlational analysis was conducted between the outcome variable and all predictors in the model leading up to Exam 1. Based on descriptive statistics and evaluation of assumptions, while some subscales were non-normal on each category of the outcome variables with considerable skewness and kurtosis present in subscales at certain time points, Levene's test for homogeneity of variances resulted in p-values>0.05 for almost all subscales at each time point indicating that the variances were equal (Tabachnick & Fidell, 2001); analysis was conducted under these distributional limitations and without transformation of any predictor variables.

Using the correlation matrix shown in **Table 6.15**, evaluations similar to those in multiple regression were conducted to find predictors that showed strong, significant correlations with the outcome variable (performance group) and relatively weaker correlations among one other. The strongest significant correlation was observed between TP total score and performance group while SE related to assessment (before exam 1) showed the strongest correlation with the outcome variable. While other pre-exam 1 SE subscales also showed moderate to strong correlations with the dependent variable, these subscales displayed significantly stronger and higher correlations among themselves; thus, pre-exam 1 SE related to assessment was the only subscale that was utilized in the model.

Table 6.15. Point-biserial correlations between predictor variables and performance group on Exam 1 for GCI (AY14-15). Higher SE/OE scores indicate lower SE/OE.

		Performance group (0 or 1) based on z-score (Exam 1)	ACT_Comp	ACT_Math	ACT_Science	SE_assess_START	SE_interpersonal_START	SE_everyday_START	SE_low_order_START	SE_strategies_START	OE_career_START	OE_performance_START	OE_learner_START	OE_lab_success_START	OE_understanding_START	TP_Math	TP_Chem	TP_total	SE_assess_PEX1	SE_interpersonal_PEX1	SE_everyday_PEX1	SE_low_order_PEX1	SE_strategies_PEX1	OE_career_PEX1	OE_performance_PEX1	OE_learner_PEX1	OE_lab_success_PEX1	OE_understanding_PEX1
Performance group (0 or 1) based on z-score (Exam 1)	Pearson Correlation	1																										
ACT_Comp	Pearson Correlation	.254**	1																									
ACT_Math	Pearson Correlation	.230**	.792**	1																								
ACT_Science	Pearson Correlation	.270**	.825**	.643**	1																							
SE_assess_START	Pearson Correlation	-.191**	-0.102	-0.086	-0.073	1																						
SE_interpersonal_START	Pearson Correlation	-0.097	-0.084	-0.071	-0.098	.635**	1																					
SE_everyday_START	Pearson Correlation	-.160*	-.138*	-.158*	-0.097	.247**	.224**	1																				
SE_low_order_START	Pearson Correlation	-0.042	0.083	0.085	0.066	.303**	.193**	.361**	1																			
SE_strategies_START	Pearson Correlation	-0.104	-.152*	-0.098	-0.089	.410**	.258**	.427**	.430**	1																		
OE_career_START	Pearson Correlation	-0.042	-0.049	-0.019	-0.032	.315**	.243**	0.077	0.038	.165**	1																	
OE_performance_START	Pearson Correlation	-0.102	-.142*	-0.104	-.167**	.366**	.262**	.171**	.149*	.193**	.484**	1																
OE_learner_START	Pearson Correlation	-0.012	-.163**	-.140*	-0.096	-0.093	-0.013	-0.103	-.187**	-0.072	-0.036	0.001	1															
OE_lab_success_START	Pearson Correlation	-0.063	-0.118	-0.117	-.134*	.313**	.179**	0.110	.130*	0.101	.510**	.567**	-0.033	1														
OE_understanding_START	Pearson Correlation	-0.086	-0.113	-0.092	-.129*	.371**	.214**	.163**	.140*	.199**	.446**	.376**	-0.054	.467**	1													
TP_Math	Pearson Correlation	.265**	.338**	.412**	.303**	-0.064	-0.040	-.154*	0.115	-0.102	0.022	-0.077	-0.003	-0.073	-0.043	1												
TP_Chem	Pearson Correlation	0.078	.248**	.286**	.252**	-0.015	0.057	-0.110	-0.054	-.138*	-0.067	-0.020	0.023	-0.032	-0.036	.323**	1											
TP_total	Pearson Correlation	.398**	.370**	.421**	.334**	-0.085	-0.003	-.270**	-0.040	-.194**	-0.014	-0.068	0.021	-0.084	-0.037	.718**	.697**	1										
SE_assess_PEX1	Pearson Correlation	-.274**	-.168**	-.212**	-.196**	.545**	.262**	.251**	.240**	.264**	0.118	.255**	-.163**	.139*	.135*	-.254**	-0.071	-.268**	1									
SE_interpersonal_PEX1	Pearson Correlation	-.180**	-0.082	-.167**	-0.093	.381**	.371**	.227**	.288**	.246**	0.052	.166**	-0.044	0.095	.138*	-.127*	0.077	-0.112	.645**	1								
SE_everyday_PEX1	Pearson Correlation	-.150*	-.224**	-.219**	-.151*	.309**	.196**	.396**	.316**	.365**	.125*	.232**	0.064	.135*	.153*	-.210**	-.168**	-.276**	.416**	.329**	1							
SE_low_order_PEX1	Pearson Correlation	-0.039	0.020	0.021	0.011	.189**	0.010	.164**	.416**	.241**	0.072	0.101	-0.012	0.050	.175**	-0.018	-.147*	-0.117	.378**	.369**	.441**	1						
SE_strategies_PEX1	Pearson Correlation	-.183**	-.148*	-.216**	-0.118	.383**	.253**	.236**	.268**	.384**	.147*	.189**	0.069	0.057	0.081	-.246**	-0.096	-.218**	.544**	.436**	.537**	.497**	1					
OE_career_PEX1	Pearson Correlation	-0.041	-.174**	-.197**	-.150*	.228**	.184**	0.101	0.036	0.116	.446**	.286**	0.107	.278**	.277**	-.140*	-.135*	-.137*	0.048	0.021	.129*	0.031	0.108	1				
OE_performance_PEX1	Pearson Correlation	-.136*	-0.092	-0.101	-0.114	.332**	.170**	.162**	.128*	.155*	.228**	.298**	-0.024	.312**	.298**	-.235**	-0.071	-.229**	.283**	.237**	.177**	.141*	.185**	.465**	1			
OE_learner_PEX1	Pearson Correlation	0.032	-0.110	-0.049	-0.092	-0.036	0.044	-0.041	-0.055	0.032	0.078	0.054	.437**	0.090	0.058	-0.006	-0.106	-0.019	-.226**	-0.091	0.005	-0.078	-0.019	.149*	0.017	1		
OE_lab_success_PEX1	Pearson Correlation	-0.031	-.175**	-.160*	-.182**	.276**	.235**	0.081	0.044	.126*	.357**	.363**	0.055	.407**	.338**	-.162**	-0.048	-.135*	.141*	.123*	0.117	0.055	.138*	.600**	.592**	.124*	1	
OE_understanding_PEX1	Pearson Correlation	-0.047	-0.055	-0.095	-0.065	.259**	.163**	.170**	.126*	0.113	.338**	.253**	-0.056	.352**	.485**	-.128*	-.140*	-.184**	0.101	0.047	.123*	.154*	0.113	.549**	.534**	0.004	.621**	1

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).
c. Listwise N=254

The size of this dataset was N=254 with 124 students (48.8%) in the low performing group and 130 (51.2%) in the high performing group. Based on the evaluations resulting from correlational analysis, two predictors (SE related to assessment – pre-exam 1 and TP total score) were used to develop a discriminant model to predict membership in performance groups on Exam 1. While the results produced were substantial, key parameters that required evaluation will be described for this model.

Univariate ANOVAs were carried out for each independent variable to determine if these variables differ for the two groups (high performing and low performing). These results are shown in **Table 6.16**.

Table 6.16. Tests of Equality of Group Means - DA model predicting group membership in Exam 1 - GCI

	Wilks' Lambda	F	df1	df2	Sig.
TP_total	0.841	47.484	1	252	0.000
SE related to assessment (pre-Ex 1)	0.925	20.387	1	252	0.000

The ANOVA results indicate that both total TP score and average score for SE related to assessment (pre-exam 1) differ (Sig. = .000) for the two performance groups. Wilks' lambda denotes the importance of the predictor to the discriminant function, with smaller values implying greater importance of the independent variable to the discriminant function.

Box's test allows for assessing the homogeneity of covariance matrices. This test presents two pieces of information, as shown in **Table 6.17**, to evaluate this assumption.

Table 6.17. Box's test of equality of covariance matrices

Box's M		8.869
F	Approx.	2.931
	df1	3
	df2	12305888.002
	Sig.	0.032

Performance group on Exam 1	Rank	Log Determinant
Low performing (0)	2	4.383
High performing (1)	2	3.679
Pooled within-groups	2	4.058

The log determinants describe the extent to which a certain group's covariance matrix differs, with larger log determinants indicating greater differentiation. Rank refers to the number of independent variables (2) in this model. As DA assumes homogeneity of covariance matrices between groups, determinants that are relatively equal would be preferred.

Box's M test evaluates the assumption of homogeneity of covariance matrices. In this model, a significant value of 0.032 indicates that groups do differ in their covariance matrices, potentially violating an assumption of DA. As outliers and transformations were not considered in this model, this was an expected result and analysis proceeded under these limitations.

Information about the discriminant function was provided by examining the eigenvalues and canonical correlations as shown in **Table 6.18.**

Table 6.18. Summary of canonical discriminant functions for membership in performance groups – GC I Exam 1

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	0.232	100.0	100.0	0.434
Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1	0.811	52.448	2	0.000

As there are only two groups in the outcome variable, only one discriminant function is produced. The canonical correlation is the measure of association between the discriminant function and the dependent variable. Squaring the canonical correlation coefficient results in the percentage of variance explained in the dependent variable. Thus, in this model, the DA suggests that two variables considered as a set are related to performance groups and explain around 19% of its variance. Conversely, the value of Wilks' Lambda denotes the amount of variance unaccounted for by this model. Thus, 81.1% of the variance in performance group membership on Exam 1 is unexplained by this model. A significant Wilks' Lambda suggests that the model is a good fit for the data.

The unstandardized discriminant function coefficients (similar to unstandardized regression coefficients in multiple regression) are used to construct the actual prediction equation which can be used to classify new cases. These coefficients are shown in **Table 6.19.**

Table 6.19. Unstandardized canonical discriminant function coefficients

	Function 1
TP_total	0.095
SE_assessment - pre Ex 1	-0.493
(Constant)	-5.391

Thus, for this model, the discriminant function was given by:

$$D_{exam\ 1} = -5.391 - 0.493\ (\text{SE assessment} - \text{pre-exam 1}) + 0.095\ (\text{TP total})$$

If variable means (rather than individual values for each student) were to be entered into the discriminant function, the resulting average discriminant scores for cases in the two groups would be referred to as centroids. In this model, the average score for the high performing group on the discriminant function was .469 while the average score for the low performing group was -.492. Using this information, if a student's score on the discriminant function was closer to -.492, then the data probably came from the low performing group. Practically, a cutting point of halfway between the two centroids would be used to determine group membership of a student. In this case, the cutting point would be equal to (-.492 + .469)/2 = -0.0115. Thus, if a student's score on the discriminant function was above -0.0115, the student probably belonged to the high performing group; if the score were below -0.0115, student was probably in the low performing group.

Lastly, the classification results shown in **Table 6.20** were used to assess the efficacy of the discriminant function in classifying students correctly and incorrectly.

Table 6.20. Classification statistics showing correct and incorrect classification of cases – Exam 1

Performance group (high or low) on Exam 1			Predicted Group Membership		Total
			Low performing	High performing	
Original	Count	Low performing	86	38	124
		High performing	36	94	130
	%	Low performing	**69.4**	30.6	100.0
		High performing	27.7	**72.3**	100.0
Cross-validated	Count	Low performing	85	39	124
		High performing	37	93	130
	%	Low performing	**68.5**	31.5	100.0
		High performing	28.5	**71.5**	100.0

* 70.9% of original grouped cases correctly classified.

** 70.1% of cross-validated grouped cases correctly classified.

Based on this table, the discriminant function correctly classified 70% of the cases, making about the same proportion of mistakes for both categories. In essence, 20% of cases were classified above chance level, which was adequate. 69% of students in the low performing group were correctly classified while 30.6% were incorrectly classified. In the high performing group, 72.3% of students were correctly classified while 27.7% were incorrectly classified. Thus, the function seems to perform equally for both groups, although the classification results are likely more reliable for students in the high performing group. The cross-validation method, called 'leave-one-out classification', classifies each case by the functions derived from all cases other than that case.

Model 2: Predicting membership in high vs. low performing groups on Exam 2

Using the same process as for Exam 1, the correlation matrix was evaluated for potential predictors; the dataset had 310 cases, with 156 (50.3%) students in the low performing group and 154 students (49.7%) in the high performing group. The average scores on Exam 1 for the high and low performing groups were 82.5 and 66.4 respectively. In the model for predicting membership in performance groups for Exam 2, two predictors were used in the development of the model: Exam 1 score and SE related to applying chemistry to everyday tasks (Pre-exam 2). The test of equality of group means was significant for both variables, with score on exam 1 being the more important variable to the discriminant function (Wilks' Lambda = .702, $F(1,308)=130.773$, p=.000).

Box's M test for equality of covariance matrices was significant (p=.046), indicating that the groups differed in their covariance matrices. The canonical correlation was .557, implying that 31% ($.557^2$) of variance was explained in the dependent variable for model 2. The discriminant function for this model was given by:

$D_{exam\ 2}$ = -5.068 - (.372 SE-applying chem. to everyday tasks_Pre-exam 2) + (.076 Exam 1 score)

Classification results for this model showed that overall, 73.2% of cases were correctly classified. The function was likely more reliable for the high performing student group as 76.6% of students in this group were correctly classified while 69.9% of students in the low performing group were correctly classified.

Model 3: Predicting membership in high vs. low performing groups on Exam 3

This dataset had 298 cases, with 148 (49.7%) students in the low performing group and 150 students (50.3%) in the high performing group. The average scores on Exam 2 for the high and low performing groups were 76.0 and 59.1 respectively. In the model for predicting membership in performance groups for Exam 3, two predictors were used in the development of the model: Exam 2 score and SE related to interpersonal tasks (Post-exam 2). The test of equality of group means was significant for both variables, with score on exam 2 being the more important variable to the discriminant function (Wilks' Lambda = .715, $F(1,296)=118.243$, p=.000).

Box's M test for equality of covariance matrices was not significant (p=.123), indicating equality of covariance matrices. The canonical correlation was .554, implying that 30.7% ($.554^2$) of variance was explained in the dependent variable for model 3. The discriminant function for this model was given by:

$D_{exam\ 3}$ = -3.675 - (.316 SE-interpersonal tasks_Post-exam 2) + (.067 Exam 2 score)

Classification results for this model showed that overall, 72.1% of cases were correctly classified. The function was only slightly more reliable for the high performing student group as 73.3% of students in this group were correctly classified while 70.9% of students in the low performing group were correctly classified.

Limitations

One of the glaring limitations of this study was not utilizing incomplete data sets to examine a true growth model and the interaction of variables in this model at each time point. It is possible that the differences observed between models could have been a product of missing data. The implementation of modeling techniques such as hierarchical or multilevel modeling, which allow for inclusion of incomplete data sets, would have enabled changes in affective variables to be tracked across a semester, both overall and by gender subgroups.

The importance of predictors, to develop the regression models in this study, was based on the bivariate or partial correlation coefficients between predictors and the outcome variables. While these methods resulted in residuals that were normal and whose averages were close to zero, there is literature that suggests methods other than correlational analysis would offer more intuitive measures of predictor importance. Methods such as dominance analysis, that rely on variance decomposition and changes in model fit, would allow for comparisons between several subset models and evaluation of predictor contributions across these models (Azen & Budescu, 2003; Johnson & LeBreton, 2004).

Although CFA was not conducted on the subset instrument, similar to related studies in the literature, the items on the subset were obtained from confirmed factor structures of the full-length surveys; in addition, the use of more than one item in the subset instrument to represent an original subscale allowed for a more thorough representation of the construct while tracking changes.

While this study monitored and assessed statistics related to residuals in a compartmentalized multiple regression analyses for each performance event, the potential of non-independence of residuals was prevalent due to time effects; if observations were obtained at different times, those from points that are closer in time would be more highly related (than those from later/earlier times); the same phenomenon would also be observed in the case of repeated measurement over time. A time series analysis or modeling techniques would have helped to remedy this situation.

While discriminant analysis is a unique method to predict group membership and build predictive profiles for students, one of the more fundamental requirements of this method is that the groups that constitute the dependent variable should be naturally occurring and mutually exclusive. Although high and low performing groups exist naturally, this study created groups by manually dichotomizing interval data into z-scores and assigning a group based on these scores. Utilizing DA required the fulfilment of several assumptions, only some of which were considered in this study. Thus, it would be irresponsible to attest to the generalizability of the results obtained using DA.

Conclusions and implications

As changes in affective measures could potentially impact performance in chemistry gateway courses, the development of predictive models to elucidate the point at which affective measures decline would be valuable in implementing targeted interventions to offset the decline in these measures. To that end, the purpose of this study was to utilize 'control' data sets in developing predictive models whose utility (model residuals) would be examined on data sets comprising of students who have participated in an intervention. As the number of students with complete survey responses was low and did not allow for examination of a time series analysis, data were examined using all relevant predictors leading up to key performance events (hourly exams) during the semester. Regression models developed for predicting scores on each hourly showed zero residual averages and normality of residuals.

This same approach was used to develop predictive models that would be able to categorize students into low and high performing groups on each hourly exam. While these models were consistently better at classifying high performing students, the low performing group did not fare too poorly in being classified correctly. Examining and identifying students at-risk for low performance and ultimately perhaps low persistence would be highly valuable in assembling persistence profiles that can be used to identify key factors that place students at higher risk for lack of persistence. It should be noted that the categorization of a 'low performing' group is from a normative reference (e.g. below average) in this case and could have demonstrated considerable proficiency on the exam (e.g. if the entire class did well).

The results of these analyses – an average of an increase in 1 point on SE/OE (affective measure) decreased the score on an exam by an average of 2.75 points, while controlling for prior academic proficiency – reveal the importance and contributions of affective and performance indicators in understanding and explaining, on a finer scale, events that could potentially trigger lower measures of persistence. The identification of these triggers and the measures they impact, whether performance, self-efficacy or outcome expectations, present an opportunity to intervene and offset the decline in these factors. Ultimately, if these interventions could trigger a change in decision-making regarding choice of major, it would immensely aid in keeping students on track for graduating with a STEM major.

CHAPTER 7: DESIGN AND TESTING OF INTERVENTIONS IN GENERAL CHEMISTRY

This chapter details the development and subsequent testing of interventions for students enrolled in GC I. Results obtained from usability studies and the implementation of these interventions will also be described.

Background and Rationale

Several efforts have been dedicated towards improving student persistence in courses and performance in specific tasks and the entire course. Institutional efforts to improve these measures have mostly involved curricular changes, the addition of remedial courses to accommodate for deficiencies, assistance with major selections and career advising. However, researchers have for some time now, initiated and sustained targeted interventions towards improving students' behavioral attributes (Toven-Lindsey et al., 2015; Pajares, 1997; Margolis & Mccabe, 2006). The introduction of social cognitive theory brought into focus several interrelated constructs which have influenced a student's choice of activities, persistence and effort. Of these constructs, self-efficacy has been the most prevalently studied factor due to its utility in being an effective predictor of learning and student performance. While no single process can explain the complexity and variations in students' motivational beliefs and efforts to learn, studies have shown that self-efficacy beliefs provide students with a sense of support that helps motivate their learning through the use of self-regulatory processes which are systematically oriented toward attainment of one's own goals (Zimmerman, 2000).

Self-regulated learners engage in self-evaluation and are active participants in their own learning (Ablard & Lipschultz, 1998). They also possess a variety of cognitive and metacognitive strategies that are employed when needed to accomplish academic tasks. One study (Zimmerman,

1999) recognized five important aspects of students' efforts to self-regulate their learning: Setting goals, context adaptations, using strategies, social processes and self-monitoring.

Research in self-regulated learning has shown that these learners are typically high-achievers; students scoring in the top 1% on an achievement test utilized certain self-regulated learning strategies more frequently. These strategies were geared toward optimizing, organizing, transforming information and providing their own rewards and punishments based on performance (Zimmerman & Martinez-Pons, 1990). Perceived verbal and mathematical efficacy and strategy were measured with fifth, eighth and eleventh grade students, and showed a 16-18% shared variance between efficacy beliefs and strategy use across the three grade levels of schooling (Zimmerman & Martinez-Pons, 1990). While trait measures of self-perceptions are fairly stable across time and setting, self-efficacy is cyclical and has a tendency to respond to changes in personal context and outcomes, regardless of the source of the efficacy beliefs (performance accomplishments, vicarious experiences, verbal persuasion and physiological states). This sensitivity allows evaluations of self-efficacy beliefs as indicators of change during instructional interventions as well as signs of initial individual differences (Zimmerman, 2000).

Studies have shown that improvements in self-efficacy have been facilitated by training students with learning and motivational deficiencies to model specific self-regulatory techniques, describe the impact of the technique and provide feedback regarding their impact (Zimmerman, 2000). The frequency and immediacy of feedback created higher perceptions of personal efficacy (Schunk, 1983). Students' attribution of feedback to their effort allowed for perceptions of greater progress, sustained higher motivation and greater efficacy for further learning (Schunk, 1987). These studies not only indicated the sensitivity of efficacy beliefs to instructional interventions, but also the mediational role of these beliefs in explaining self-regulation and achievement outcomes in learners (Zimmerman, 2000).

Given the importance of self-efficacy beliefs in playing a causal role in students' academic development, use of learning strategies and persistence, the goal of this study was to design interventions which could positively impact student performance, self-efficacy or outcome expectations, and provide a better understanding of how to maximize the likelihood of keeping students on track for graduating with a STEM major.

Using results from the subset instrument and the hypothesis that changes in affective measures occur prior to and following important performance events, this study was an opportunity to design study tools (intervention) which could ultimately target students with low performance or low affective measures (perhaps at a critical time point) and offset the decline before these measures resulted in a lack of persistence. Two objectives guided this study:

a) To examine changes in predicted performance and affective measures in students who had completed the intervention.
b) To evaluate student usability of these interventions, especially their engagement with problem solving strategies.

Methodology

This section describes the process used in designing the intervention. The format, construction of the intervention module, the testing and the participants will also be detailed. Analyses conducted to evaluate the efficacy and usability of the interventions will be described.

Design of the intervention module

The interventions developed for this study were aimed at targeting one, two or all three measures examined during development of the predictive persistence model: performance, self-efficacy and outcome expectations. The occurrence of a performance (testing) event was likely to trigger changes in students' affective measures, especially following a testing event. For instance, if students' experienced lower self-efficacy following a testing event in GC I, the goal was to target this lower self-efficacy and increase confidence by planning course specific tasks tied to the content area that perhaps triggered the decline. Thus, these tasks had to be representative of instructional material in the course; in addition, consideration also had to be given to the level of challenge associated with the tasks so as to not artificially inflate or lower confidence or doubt students' abilities (Pintrich & Schunk, 1995). Consequently, instructional-level tasks had to be slightly above the student's current performance level (Margolis and Mccabe, 2006). In order to avoid issues with selecting relevant and overly challenging or simple tasks, this intervention tool utilized problems from past exams in the course to examine changes in self-efficacy. These problems were a combination of multiple choice and free response items, similar to the format of the hourly exams.

The second component to these tasks was the feedback offered to the student upon completion of a problem. As there was no score assigned to the problems in this intervention module, there was less importance placed on the correctness of the answer. Students were informed of the correct answer (regardless of the correctness of their answers), commended on their effort for a correct answer with a nod to the difficulty of the problem if the subject matter was one with which students usually struggled e.g. "Great job on getting this question correct. This was not an easy problem!". While this was the extent of feedback in the S16 experiment (the first implementation of the intervention), students' comments indicated that 'knowing the answer alone' was of little benefit to them without a solution to which they could compare their work. Consequently, in F16, following the assessment about the correctness of their answer, students were offered a detailed explanation of the solution to the problem. This instructive feedback was given to aid students, especially struggling learners, who perhaps benefited from a comprehensive map to compare and correct their mistakes, thereby correcting their understanding of material.

The third component of this module was the inclusion of problem solving strategies specific to the problem being solved; in addition, feedback was offered to students based on their selection of strategies. The type of strategies to include was dependent on the specific type of problem solving. As general problem solving involves four major steps – understanding the problem, devising a plan to solve the problem, implementing a solution plan and reflecting on the problem (Polya, 1957) – and each step involves smaller tasks, strategies for each problem were also divided based on what students did 'at the start of the problem', 'during the problem' and 'as they finished the problem'. Within these three categories, task specific strategies were listed based on common ways in which students approached problem solving in chemistry (Bodner & Herron, 2002). For example, a GC I student attempting to solve a problem related to calculating density using water displacement would be shown the strategies as displayed in **Table 7.1**. While some of these strategies were applicable to most quantitative problems, others had to be customized for the problem at hand. Using **Table 7.1** as an example, 'recalling conversion factors or equations' to start the problem and 'using dimensional analysis or diagramming the scenario' while solving the problem were specific to the question posed.

Table 7.1. Example of problem solving strategies contained in intervention module – GC I

What was the most important strategy you used to **start** the problem?

- ◯ Identifying and writing down the information given and asked for in the problem
- ◯ Recalling conversion factors to link the given unit to the final unit
- ◯ Recalling a rule, equation or principle to solve the problem
- ◯ Finding a similar problem and plugging in the numbers as shown in that problem
- ◯ Recalling a similar question done in lecture/practice exams/textbook/discussion
- ◯ Some other strategy not listed here. Please provide:

What was the most important strategy you used **while** solving the problem?

- ◯ Using dimensional analysis to solve the problem
- ◯ Performing the mathematical calculations
- ◯ Breaking the problem down into a series of steps that will enable me to determine what I need to know
- ◯ Diagramming the scenario or system either macroscopically or on the particle level
- ◯ Finding a similar problem and plugging in the numbers as shown in that problem
- ◯ Did not write anything down; starting plugging in numbers in my calculator
- ◯ Some other strategy not listed here. Please provide:

What was the most important strategy you used as you **finished** the problem?

- ◯ Applying the correct number of significant figures to the final answer
- ◯ Reviewing: re-reading the original problem and selecting the correct answer
- ◯ Guessing
- ◯ Did not write anything down; plugged in numbers and conversion factors directly my calculator to get the answer
- ◯ Some other strategy not listed here. Please provide:

After considering how you did on this problem and the strategies you used to solve this problem, how will this change how you study for this material and prepare for your test?

The fourth component of this module (to complement the strategies themselves) was the feedback students received upon selection of a problem-solving strategy. The feedback was proposed by researchers (and loaded into Qualtrics) depending on the strategy selected by students. This feedback was provided to not only commend students for effort, and persistence in working on moderately challenging tasks, but also to offer students a logical and systematic sequence of steps for approaching difficult tasks. Commending students on the use of effective strategies was expected to reinforce their beliefs of their own cognitive and metacognitive strategies. In addition, the identification of ineffective strategies was expected to engage students' metacognitive strategies with the expectation that this could change. The ultimate goal of this approach was, of course, to impact students' self-efficacy and performance.

While outcome expectations and self-efficacy are correlated to a moderate degree, with commonality between some subscales in the COES and CSEAS, interventions aimed at targeting low outcome expectations were not designed to be as task specific as SE as some dimensions of OE involved proximal and distal outcomes. Instead, targeting low OE necessitated an understanding of students' expectations, their goals and subsequently helping them establish a connection between their actions and expectancies; students also reflected on how they expected to prepare for their exams, the resources that would be utilized and whether the feedback on their performance and strategies had made them aware of their learning.

The entire module was designed on the Qualtrics platform; as Qualtrics allowed for different question types, some multiple-choice questions, such as selecting a figure, were displayed using a 'hot spot', which represented the figures and accompanying letter choices in a colored region; students had to click on a region vs. a radio button to indicate their answer. Each question was validated to be 'forced response', with open response items allowing a maximum of 100 characters. Students who responded correctly to a problem could view feedback on their answer and a detailed explanation on the same page as the question. Students who answered questions incorrectly viewed the question, their chosen answer, the correct answer and explanation on the subsequent page; this was done to dissuade students from changing their answers after viewing the explanation on the same page. Students were introduced to the module using a cover page which offered a brief description of the module and its constituent tasks, in addition to requesting students' first and last names. The module for each hourly exam followed the same format; to stay consistent with material covered in class and reflect coverage on the upcoming exam, questions differed in content and in some cases, quantity.

Testing interventions – Classwide usability study

These interventions were only tested with students in GC I; although the intention of these interventions was to target students at-risk, the intervention was provided to the entire class. As it is expected that there would be multiple opportunities for interventions as needed throughout one semester, these interventions were packaged as "study packets/tools" for students to complete, thus integrating the necessity to complete these into the course.

Testing the interventions involved studying the usability of these interventions using a hybrid of eye-tracking studies and semi-structured interviews. Eye-tracking is a technique that captures eye behavior in response to a visual stimulus such as a computer interface, photograph or page in a newspaper. An eye tracker captures eye movements and determines the position of one or both eyes multiple times per second. Fixation duration is a brief glance lasting between 100-300 milliseconds, although longer fixation ranges have been documented by some researchers (Palmer, 2002). Usability testing is a technique that evaluates the ease of use of a product. In usability testing, a representative sample of actual or potential end users is asked to attempt real tasks using the product (Dumas & Redish, 1999). Combining eye tracking with usability testing allows for a more complete picture of the interaction as eye movement analysis can support findings based on behavioral measures (Dumas & Redish, 1999).

In this study, as the emphasis in the module was on selection of effective problem solving strategies, it was crucial to examine students' engagement and interactivity with the pages that displayed these strategies. Thus, each page of the intervention module was coded (using source code from Qualtrics) into a html webpage (stimulus) and sequentially loaded into the SMI RED500 remote eye-tracking system, used in conjunction with the software SMI iView NG, SMI Experiment Center v3.7, and SMI BeGaze v3.7.40. As the intention was to track eye movements on the computer screen and more importantly aggregate results between users, the stimulus had to be a static page without scrolling capabilities. Consequently, feedback could not be displayed on the same page as the question and problem solving strategies corresponding to each phase of problem solving had to displayed on separate pages. In this study, each trial was a page that displayed a problem-solving strategy; as there were eleven questions in the study packet, there were 33 trials related to the problem-solving strategies.

GC I students were solicited during the last five minutes of lecture a week before their second hourly exam; the second exam was selected due to its importance in being a key point following which students make decisions about staying in or dropping the course (drop date for the course was after exam 2). Sixteen students signed up to participate in the usability study for which compensation was an ACS study guide; the process lasted between 45-60min. Most students were biochemistry and biology majors, with a select few on a pre-med or pre-PT (physical therapy) track. The study packet that was loaded into the eye tracking system was designed for exam 2.

At the start of the interview, students were provided instructions on the experiment process. This was followed by instructions on how to use the equipment after which a nine-point eye tracking "calibration" (sampling rate 60 Hz) was conducted by showing dots at several baseline positions on the screen (corners, center) and having the student fixate on them. Students were instructed to get situated comfortably in a position that would minimize movement during the experiment. As less emphasis was placed on the way students solved a problem as opposed to how they interacted with the strategies, students were instructed to approach the problems as they would if they worked through the study packet on their own time. The course textbook and scratch paper were available for use as was the interviewer to answer any content specific questions such as providing an equation or offering a page number in the textbook for students to quickly access relevant tables.

In order to minimize movement between the screen and scratch paper while solving a problem, students were asked to focus on the problem, record all relevant information and subsequently utilize a binder for reviewing or re-reading the problem; this binder contained paper copies of all the problems as they appeared on the computer screen. Students were instructed to solve the problem, verbalize their answer to the interviewer to check for correctness, following which students could access the detailed explanation of the solution on the screen. As stated before, whether or not students accessed feedback was of less interest than was how they approached the selection of the strategies used in each problem. Moreover, as several students had not started studying for the exam at the time of the interviews, they required considerable help and did not know how to solve some of the problems, especially those covering content that was pending instruction in class.

For the pages containing problem solving strategies, the independent variables were the stimuli themselves, while the dependent variables were measured as students' reactions to those stimuli. Information examined for these pages included time on task, fixation times and counts, scan paths, pupil diameter and time on areas of interest (AOIs). A scan path is a repetitive succession of eye fixations (Brandt & Stark, 1997). Josephson and Holmes (2008) conducted studies in which they found that participants had preferred scan paths and that different participants exhibited similarities in eye movement sequences. Areas that were designated as AOIs were the question itself ("what was the most important strategy you used **while** solving the problem"), the bold and underlined term denoting the stage of problem solving, the choices as a whole and the words at the start of the choices as these terms characterized actions that could perhaps indicate behaviors of students with lower affective measures. Data collected (fixation times, counts, scan paths) were exported into Excel for analysis. As the outcome expectations component of the study packet consisted mainly of student reflections and expectations, these questions were asked of students during the interviews as opposed to having them type answers on a static page.

Using eye tracking measurements, studies have shown that fixation count, duration and average fixation rates on particular locations are indicative of visual attention, which triggers mental processes to solve a given task (Just & Carpenter, 1980). Changes in pupil dilation are also indications of cognitive workload (Laeng et al., 2012). While there was no empirical data for the actual problem solving process as students' eyes were off screen and thinking aloud was minimal on account of students trying to find information to solve a problem or struggling to solve the problem, it was hypothesized that measures such as fixation times, counts and time spent on certain regions on the pages containing strategies would offer evidence for how useful students found the feedback and strategies.

Data collection and participants

The intervention ("study packet") has been in use since Spring 2016; data from Spring 2016 – Fall 2016 were used as the intervention dataset while data from Fall 2014 – Fall 2015, (collected using the subset instrument described in chapter 6) served as the control dataset. The studies described in this chapter were conducted at a large, urban, research intensive public university in the Midwestern United States.

The study packet was distributed to GC I instructors using a link generated by Qualtrics. Subsequently, instructors sent out the link to their students and the study packet was made active (opened) a week before each hourly exam and data collection was stopped on the day of the exam (a few minutes before the start of the exam). All responses (complete and in progress) were downloaded to Excel but only those who had submitted complete study packets were deemed to have been exposed to the intervention. Students who did not start the intervention or were recorded as being 'in progress' were excluded from analysis. For completing the study packet and pre/post subset surveys, students were given 5 extra credit points.

As the subset models were compartmentalized for each testing event, the numbers represented in **Table 7.2** account for students who were part of these models and took the interventions. The asterisk next to Exam 3 indicates that the numbers represent intervention participants from Spring 2016 only. As data from Fall 2016 were not initially incorporated into the model, analyses pertaining to Exam 3 only include data from Spring 2016.

Table 7.2. GC I intervention participants (by gender) for three testing events. *Exam 3 shows participants for Spring 2016 only.

Spring'16 + Fall'16	Exam 1	Exam 2	**Exam 3***
Total	103	129	73
Males	30	36	19
Females	73	93	54

Data analyses

Data were cleaned as described in chapter 3. As the interventions were offered prior to a testing event to observe changes in affective measures before and after the event, descriptive statistics were obtained for subscales in the subset survey pre/post each hourly exam for assessments of univariate normality, skew and kurtosis.

Predictive validity - Standard multiple linear regression (SMLR)

In this study, the regression models developed (in chapter 6) for predicting performance on each testing event were used to predict exam scores for students who had taken the intervention. The average of the residuals was examined for changes due to the interventions. If the residual average was positive, the interventions had improved performance (positive impact) on the exam. Conversely, if the residual average was negative, the interventions had brought about a decrease in performance for those who had taken the study packet. It was anticipated that the residual averages would be slightly positive, signifying an improvement in performance.

Predicting group membership – Discriminant analysis

Using the discriminant functions developed (in chapter 6), group membership was predicted for students who had completed the intervention. Each discriminant function was used to calculate a discriminant score; if the score was above the average score calculated using centroids, the student was placed in the high performing group; if the score was lower than centroid average, students were in low performing groups. Percentages of those who had been correctly classified into high or low performing groups were noted. More importantly, misclassifications were examined for movement of students between groups. Thus, if a student had been classified (predicted) as low performing, but was actually in the high performing group, this was considered a positive movement (change); if the number of students misclassified as low but ending up in a high performing group exceeded those who had been misclassified as high but ended up in a low performing group, this was considered a positive change brought about by the intervention – similar to residual analysis in multiple regression.

Odds ratios (ORs)

To evaluate the impact of the interventions on SE and OE subscales (from the subset instrument) before and after each testing event, odds ratios (ORs) were calculated for each subscale using two datasets: control data (not exposed to the intervention – AY14-15) and treatment data (exposed to the intervention – S16-F16). Odds ratios are useful because as an effect-size statistic, they give direct information about which treatment approach has the best odds of benefiting the individual (McHugh, 2009). In this study, for both control and intervention datasets, a higher affective measure was the target (dependent) indicator, and being exposed to the intervention was an independent indicator. As these interventions were targeted toward lower measures of SE or OE, the odds ratio was calculated to determine whether the odds of moving to a higher SE group on the same subscale were better for students in the control dataset or for the student group that had been exposed to the intervention. High and low SE or OE groups were designated based on raw survey responses, where a mean score > 3 was a low affective group and mean score < 3 was a high affective group. Confidence intervals and significance statistics were also determined for each OR. An example of the setup to calculate OR (in this study) is shown in **Table 7.3**.

Table 7.3. Example of set up for calculating odds ratios (OR)

	Control (no intervention) F14-F15	Exposed to intervention (S16 + F16)	Odds
Stayed in low affective group	a	b	a / b
Moved into high affective grp.	c	d	c / d
Totals			(a*d) / (b*c)

Results and Discussion

Descriptive statistics prior to and following each testing event are displayed in **Tables 7.4** and **7.5**. These statistics correspond to students who were exposed to the interventions (S16 – F16) and are organized by each subscale of the subset instrument.

Table 7.4. Descriptive statistics for outcome expectations subscales before and after testing events (hourly exams) – Treatment (intervention) group in GC I (S16-F16)

Subset factor	Time point	N	Min	Max	Mean	Std. dev.	Skew	Kurtosis
OE - career	Pre Ex 1	103	1.00	4.33	1.74	.57	1.60	5.15
	Post Ex 1	155	1.00	4.00	1.75	.55	.69	1.49
	Pre Ex 2	155	1.00	5.00	1.80	.64	1.43	4.45
	Post Ex 2	73	1.00	3.00	1.78	.54	.05	-.67
	Pre Ex 3	73	1.00	3.67	1.78	.63	.74	.25
	Post Ex 3	72	1.00	3.00	1.75	.56	.12	-.71
OE - performance based tasks	Pre Ex 1	103	1.00	4.00	1.73	.79	.97	.48
	Post Ex 1	155	1.00	5.00	1.91	.90	1.09	1.23
	Pre Ex 2	155	1.00	5.00	1.96	.90	1.00	1.22
	Post Ex 2	73	1.00	4.00	2.01	.90	.70	-.27
	Pre Ex 3	73	1.00	5.00	2.10	.99	.89	.39
	Post Ex 3	72	1.00	5.00	1.99	.92	.91	.72
OE - learner based tasks	Pre Ex 1	103	1.00	4.00	2.38	.67	.26	-.46
	Post Ex 1	155	1.00	5.00	2.40	.84	.48	.00
	Pre Ex 2	155	1.00	4.50	2.42	.75	.40	-.03
	Post Ex 2	73	1.00	5.00	2.60	.88	.65	.49
	Pre Ex 3	73	1.00	5.00	2.66	.94	.35	-.51
	Post Ex 3	72	1.00	5.00	2.58	.78	.59	.51
OE - lab success	Pre Ex 1	103	1.00	4.00	1.93	.63	.92	2.21
	Post Ex 1	155	1.00	5.00	1.95	.68	1.18	3.52
	Pre Ex 2	155	1.00	5.00	1.89	.72	1.47	4.55
	Post Ex 2	73	1.00	3.50	1.93	.60	.17	-.14
	Pre Ex 3	73	1.00	4.00	1.97	.67	.52	.24
	Post Ex 3	72	1.00	4.00	1.90	.65	.88	1.96
OE - understanding chem.	Pre Ex 1	103	1.00	4.00	2.02	.63	.63	1.03
	Post Ex 1	155	1.00	5.00	2.02	.64	.78	2.56
	Pre Ex 2	155	1.00	5.00	2.02	.68	.90	2.32
	Post Ex 2	73	1.00	3.33	2.04	.59	.20	-.13
	Pre Ex 3	73	1.00	4.33	2.07	.68	.71	1.09
	Post Ex 3	72	1.00	4.00	1.98	.64	.40	.69

Table 7.5. Descriptive statistics for self-efficacy subscales before and after testing events (hourly exams) – Treatment (intervention) group in GC I (S16-F16)

Subset factor	Time point	N	Min	Max	Mean	Std. dev.	Skew	Kurtosis
SE - assessment	Pre Ex 1	103	1.33	5.00	2.70	.88	.64	-.16
	Post Ex 1	155	1.00	5.00	2.78	1.03	.43	-.52
	Pre Ex 2	155	1.00	5.00	2.80	1.02	.43	-.70
	Post Ex 2	73	1.00	5.00	2.93	1.03	.23	-.74
	Pre Ex 3	73	1.00	5.00	3.07	1.05	.06	-.85
	Post Ex 3	72	1.00	5.00	2.98	1.00	.25	-.73
SE - interpersonal	Pre Ex 1	103	1.00	5.00	2.55	.93	.46	-.04
	Post Ex 1	155	1.00	5.00	2.57	1.06	.72	.08
	Pre Ex 2	155	1.00	5.00	2.68	1.10	.42	-.57
	Post Ex 2	73	1.00	5.00	2.80	1.12	.23	-.58
	Pre Ex 3	73	1.00	5.00	3.04	1.14	.28	-.74
	Post Ex 3	72	1.00	5.00	2.56	1.05	.50	-.19
SE - applying chem. to everyday tasks	Pre Ex 1	103	1.00	5.00	1.87	.79	1.09	1.58
	Post Ex 1	155	1.00	4.00	1.86	.72	.88	.54
	Pre Ex 2	155	1.00	3.33	1.67	.65	.84	-.12
	Post Ex 2	73	1.00	3.33	1.64	.62	.65	-.64
	Pre Ex 3	73	1.00	4.00	1.66	.69	1.34	1.54
	Post Ex 3	72	1.00	4.00	1.55	.61	1.41	2.74
SE - low order tasks	Pre Ex 1	103	1.00	4.50	2.01	.72	.74	.67
	Post Ex 1	155	1.00	4.50	2.11	.83	.72	.15
	Pre Ex 2	155	1.00	4.50	2.07	.73	.87	1.02
	Post Ex 2	73	1.00	4.00	2.06	.75	.58	-.03
	Pre Ex 3	73	1.00	5.00	2.11	.80	.99	1.63
	Post Ex 3	72	1.00	4.50	1.97	.79	.99	1.35
SE - applying general chem. strategies	Pre Ex 1	103	1.00	4.50	2.07	.73	.80	.39
	Post Ex 1	155	1.00	5.00	2.07	.80	.81	.61
	Pre Ex 2	155	1.00	4.50	2.06	.74	.67	.37
	Post Ex 2	73	1.00	4.00	2.01	.73	.72	.65
	Pre Ex 3	73	1.00	5.00	2.32	.85	.87	1.02
	Post Ex 3	72	1.00	4.00	1.99	.72	.58	.40

Descriptive statistics indicate subscales with high kurtosis (>2) at certain time points. Although some subscales were non-normal and had high skewness and kurtosis, testing of predicting models (developed in chapter 6) using these data was carried out without any transformation to the subscales.

Standard multiple linear regression (SMLR)

The predictive model residuals, residual averages from the control dataset and regression equations for each testing event are shown in **Table 7.6**.

Table 7.6. Summary of predictive model residuals (control and treatment datasets) and regression equations – GCI (F16+S16 for exams 1 and 2; only S16 for exam 3)

Residual averages	Control (F14 - F15)	Intervention (S16 - F16)	Regression equations
Exam 1	-0.01	+ 2.10	Exam 1 score = 30.034 + (.761 TP_total) + (-3.573 SE_assessment - Pre Ex 1)
Exam 2	-0.02	-2.78	Exam 2 score = 33.109 + (.630 Exam 1 score) + (-1.903 OE_learner based tasks -3.871 SE_strategies - Pre)
Exam 3*	0.01	-13.0	Exam 3 score = 33.960 + (.608 Exam 2 score) + (-1.659 SE_interpersonal - Post Ex 2)

The residual averages for exam 1 are positive, indicating an improvement in scores as a result of the intervention. The results for exam 2 were unusual because residual analysis of the S16 intervention dataset alone gave residuals of +2.4 but with the addition of the F16 data, the residual averages were negative indicating a decline in performance for the students who had been exposed to the intervention. Student data from only the F16 treatment set were examined and several combinations of the predictor variables were attempted for which corresponding residual averages were recalculated; for instance, based on the predictors for exam 2, student combinations of low SE, low OE and low performance (using z-scores of subscale scores) were created and resulting residual averages were examined. Some of these results are summarized in **Table 7.7.**

Table 7.7. Combination of predictor variables and resulting predictive model residuals - GC I (F16)

Fall 2016 (GC I intervention dataset) - variables used to predict exam 2 performance	Residual averages
Low SE, low OE and low exam 1 performance	-7.83
Low SE, high OE and Low performance	-5.23
Low SE, high OE and high performance	2.62
Low SE, low OE and high performance	3.35
High SE, high OE and high performance	-9.56

The only combination that gave positive residual averages was by including low SE, high performance, and either high or low outcome expectations. Given that the residual averages were increasingly negative with high performance and affective measures, the possibility of one of these variables having an adverse impact on exam 2 performance despite the intervention warrants this data set be investigated more thoroughly. While the factors responsible for the highly negative residual averages for exam 3 were not investigated in this study, it is possible that with the scale of course obligations and preparation for the final exam and lab practical, students did not spend enough time on completing the packet thoroughly. Given the results for exam 3 are only from Spring'16, the data set from Fall'16 needs to be examined individually and in combination with Spring'16 to evaluate the impact of the intervention.

Predicting group membership – Discriminant analysis

The results of discriminant analyses classifications are shown in **Table 7.8**.

Table 7.8. Classification table for treatment (intervention) dataset – GC I (S16+F16 for exams 1 and 2; *only S16 for exam 3)

		Exam 1	Exam 2	Exam 3*
	N	103	129	73
	% correctly classified	68.0	76.7	79.5
	% misclassified	32.0	23.2	20.5
% movement among misclassifications	Low to high group	69.7	27.0	4.88
	High to low group	32.3	73.0	95.12

Similar to multiple regression, **Table 7.8** displays the efficacy of the discriminant function in correctly classifying students and the cost of misclassifying students. Two costs are associated with classification in discriminant analysis: The true misclassification cost per group and the expected misclassification cost per observation (Guo et al., 2007). Although exam 1 has the lowest percentage of correctly classified students, the cost of misclassifications is not problematic as the function appears to shift the classification of cases (among the misclassifications) toward the high performing group - as demonstrated by a higher percentage of students (69.7%) moving from the low to high performance group - thereby indicating an acceptable misclassification rate.

For exams 2 and 3, the cost of misclassifications is higher despite the greater percentage of correctly classified students. This is also in agreement with the results obtained from multiple regression, which indicated an improvement in performance on exam 1 but poor performance on exams 2 and 3.

As the subset instrument was administered and collected in the middle of the intervention time period (when the intervention was open on Qualtrics), it is possible that the study packets for exams 2 and 3 revealed what students don't know, perhaps increasing their anxiety and resulting in a higher mean value on SE subscales (low self-efficacy). It is also possible that the accompanying detailed solution lulled the students into complacency with an 'I know this' mindset. Furthermore, the multiple regression model was built with a different exam set; it is possible that exams 2 and 3 for the control group were sufficiently different than those for the semesters in which the interventions were implemented, resulting in a method (comparison of the intervention groups to the control) that was flawed.

Odds ratios

While multiple regression and discriminant analysis examined the impact of persistence measures on each testing event, the changes in these measures prior to and following a testing event were evaluated using odds ratios. In this study, the two groups were the control group (did not experience intervention) and the treatment group (exposed to intervention); the event (occurrence) was being retained in a low self-efficacy group. As the question was to determine the odds of students in low affective subscales improving their confidence or outcome expectations (moving to higher affective subscales), the target variable was a higher subscale score (based on raw scores greater than or less than 3, which was neutral in the survey) after each testing event. In this study, two sets of odds ratios were calculated: The first set involved calculating the odds of students moving from a low affective subscale (score > 3) to a high affective subscale (score < 3). The second set of ratios involved calculating the odds of students improving their affect from a low affective subscale (≥ 4 and ≤ 5) to a slightly higher affective subscale (> 3 and < 4).

In general, OR estimates of 1 mean that both groups/categories have the same odds and there is no association between the suggested exposure (intervention) and the outcome (staying in a low affective group). Estimates greater than 1 would indicate that the odds of exposure to the intervention are positively associated with the adverse outcome (staying in a low affective group) compared to the odds of not being exposed to the intervention. Estimates less than 1 imply suggest that odds of exposure to the intervention are negatively associated with the adverse outcome.

Confidence intervals and significance values were calculated for ORs corresponding to all subscales; both self-efficacy and outcome expectations subscales displayed non-significant ratios for all three testing events; the odds of students in a low affective subscale improving their confidence or outcome expectations on that subscale were no higher or lower for the control vs. intervention groups at all three testing events.

The non-significant results for SE subscales at all three testing points were unexpected due to the significant changes observed in SE subscales across a semester and the expectation that these changes might be observable during a semester. In addition, although not displayed here due to the non-significance of the ratios, there was no difference in control vs. intervention groups on any SE subscale even during the period after the completed hourly exams and before the next one. However, based on student interviews discussed in previous chapters, most students admitted to not having sustained declines in their confidence after their exams (even after viewing their grades); while the anticipation of a score may have resulted in temporary dips in confidence, once the students determined how to correct the mistakes on their exams, their confidence returned to their previously reported levels. Thus, whether SE measures change considerably at testing events with or without an intervention requires a more in-depth evaluation, perhaps into contextual or other behavioral factors. In addition, this study looked at students who merely completed the intervention; the nuances of how each problem in the study packet may have impacted students' SE would offer a richer assessment of self-efficacy's role on a much finer scale.

The non-significance of ratios with regards to outcome expectations subscales could indicate some problems with the way students perceived outcome expectations as it was operationalized in the study packet. It is possible that the outcome expectations component was not targeting the associated subscales as intended. The OE component in the study packet was focused on students' study practices, course specific (especially assessment related) goals and what steps students took in order to achieve these goals. As the OE component was operationalized from a much broader perspective such as course/career goals, explicit connections to targeted OE subscales were less likely to be observed. Examining and coding students' detailed responses to the OE statements in the study packet would offer more insight into the contextual nuances of students' expectations and any emergent associations with OE subscales.

Usability studies – Eye tracking and student interviews

The results of the eye tracking data collected indicate that students appear to be engaging meaningfully with the study packet. Although 16 students signed up to participate in interviews, data from 10 students were used for analyses as the eye tracking calibrations were successful only for these students.

The results detailed here have been analyzed across these trials. Average fixation duration in this study was ~212 ms, indicating that information was being discerned from a display (Poole & Ball, 2006). When examining by trials, these durations ranged from 152.9 ms to 278.8 ms indicating that students were viewing some areas longer than others. As fixation during does not always indicate positive attention and longer fixations could indicate confusion, these durations were evaluated based on the pages displaying problem solving strategies. It was determined that long average fixation durations were observed for pages which were populated with strategies or had longer statements in the choices. This duration was longer when the area being examined was closer to the stem of the answer choices than away. The average number of fixation counts was ~66, with larger counts (~300) demonstrated by 1-2 participants; when the interview and scan paths for these participants were assessed, the high count was likely a byproduct of students trying to find their way around the page or students fixating on an area while answering a question asked by the interviewer. The number of fixations across a page were considerable either at the stem of the statement (with a focus on the verb – "calculating, performing") or in the case of some participants, on the bold and underlined word describing the problem-solving phase (start, while, finished). When examining dwell times by areas of interest, higher dwell times were observed at the stem of the question (~6581.4, 1866.2, 1449.8 ms) and answer choices with these times decreasing in areas that were further away from the question (433.2, 699.9, 416.3 ms).

The scan paths obtained in this study were fairly varied across trials with a few key features: Students either focused on the body of the page especially when there were several choices and the density of material on the page was substantial or toward the left of the page next to the radio buttons when the selections were short statements.

These results indicate that students appear to be engaging meaningfully with the study packet. While there was minimal insight into their problem-solving process, almost all students had a consistent group of strategies that were used regardless of their performance on the problem. As some students were not thoroughly prepared at the time of the interview, they selected 'guessing' as a problem-solving strategy; however, this was always selected in combination with strategies that were part of students' normal problem solving process such as writing down information or reviewing the solution before selecting an answer. Students mentioned that the only time strategies such as 'recalling a similar problem done in lecture' would be selected is if they were at a complete loss on how to approach a problem or it was a complex multi-step problem, in which case some students were inclined to memorize the series of steps. Based on the interviews and open response items in the study packet, it appeared that while most students were appreciative of explanations to solutions and the study packet aided in their planning of material that needed to be reviewed, the strategies themselves had minimal impact in guiding students toward efficient problem solving.

Limitations

While this study examined the impact of interventions on affective and performance measures, exploring these changes thoroughly, by evaluating the strategies that students selected when working through the study packet, was not examined. Assessing the strategies used for each problem and the subsequent change in subscales could have allowed for a more in-depth analyses and better understanding of study packet's impact on performance and persistence measures.

Students were recruited for interviews a week before their scheduled hourly exam. While some students had just started to attempt the study packet as part of their preparation for the exam, there were others who had not started preparing, were unfamiliar with the content and generally struggling to solve some of the multi-step numerical problems. These students did not verbalize their thought processes and guessed most of their answers. As evidence for how useful students found the feedback and strategies (and how it was used) was based on coordinating the eye-tracking results with student articulations, students without a walk-through of their process did not provide a complete picture of how useful strategies and feedback were or how they were used.

Additionally, the eye tracking data only focused on the interactivity of the students with problem solving strategies in the study packet; the study would have been well complemented by having the students verbalize their problem-solving process. However, given that this intervention was packaged as a study tool and students had multiple opportunities to attempt it, students were quick to guess in an attempt to move on to the next problem. Moreover, one or two students found the packet useful but wanted to bypass the 'pesky strategy' pages. Thus, despite engaging meaningfully and operationalizing these strategies effectively, it is difficult to ascertain if the study packet actually improved students' problem solving strategies or was mainly used as practice problems with detailed solutions.

As with most self-selection measures, there is a strong possibility of bias associated with self-selection into the intervention.

Conclusions and implications

The use of targeted interventions to improve persistence for students with low performance or affective measures is essential in offsetting students' decisions to drop a course or change out of a STEM major. Using these interventions to influence performance or persistence measures is likely to have some impact on a student's decision making process about their intended majors. To that end, the purpose of this study was to evaluate the use of interventions by testing their impact on students' performance and affective measures prior to and following a testing event.

Using the predictive models developed in chapter 6, this study examined the utility of these models by testing them on GC I datasets comprising of students who had been exposed to the intervention; packaged as a study packet and integrated into the course, this intervention was offered to all students and could be attempted multiple times as needed before the upcoming exam.

Performance changes were examined using predictive model residuals from the multiple regression equations developed for each testing event. The results for exam 2 were troubling due to negative model residuals after students' exposure to the intervention. While it is possible that the affective measures in this model might not have been as impactful for this treatment group, these results necessitate a deeper understanding of the predictors involved or a refinement of the model itself, especially because this testing event serves as a crucial decision making point for students to stay in or drop out of the course. These results also called into question the degree to which students were engaging with the intervention and the problem-solving strategies in particular.

While eye tracking results offer a sense of student interactivity with the strategies provided in the study packet, interviews and further probing of interventions are essential to understand the processes that have the most impact on students' affective measures, performance and in a much broader context, their decision-making process about persisting in their intended STEM majors. The empirical models developed and tested in this study to examine changes in affective measures and performance are among the preliminary steps to identify the points at which affective measures decline, the triggers responsible for lowering these measures and the possible ways to offset the decreased affective components.

CHAPTER 8: DEVELOPMENT OF PERFORMANCE AND STEM PERSISTENCE MODELS IN GENERAL CHEMISTRY GATEWAY COURSES

This chapter will discuss the methodology used to develop a model integrating persistence and performance indicators. The research design, sample, data analysis and limitations of the model will also be described.

Introduction

The development of persistence models requires the integration of performance and persistence measures – self-efficacy and outcome expectations. Combining these measures in a validated persistence model will allow for the best identification of at-risk students based on where a lack of persistence occurs and what component of the model shows a deficiency. Using the definition of persistence as an "individual phenomenon", which describes students' intentions to "persist to a goal" (Reason, 2009), these goals being completion of courses or completion of degrees (Reason, 2009), this study was conceptualized from three perspectives:

a) Course performance: Using pre-affective and cognitive measures as predictors, 'local' performance models were developed to examine variables that were most influential in predicting course (or content) performance outcomes, measured by score in the course or on the final exam. Changes in affective measures (self-efficacy and outcome expectations) could indicate students who are at risk due to gateway course performance.

b) Course persistence: This was perceived as sustained enrolment and completion of a course (for example, GC I) or enrolment in the sequential course (GCII). As an outcome, this would involve recording whether a student stayed through GCI (cross-sectional) enrolled in GCII or stayed through GCII (longitudinal).

c) Persistence in a STEM major: This question was addressed by tracking changes in students' self-reported majors at the start and end of GCI (cross-sectional) or at the end of GC II (longitudinal). Using pre-affective and cognitive measures as predictors, a STEM persistence model was developed to examine variables that were most influential in predicting whether a student stayed in his or her intended STEM major (outcome) within the context of general chemistry. The development and subsequent testing of this model overall and by subgroup would be useful in highlighting differential persistence for underrepresented students, particularly female students.

Using the SCCT model of career choice as a conceptual framework, models of performance and persistence were developed in this study using general chemistry courses that constitute the two-semester sequence of gateway courses. Given that gateway courses in physical sciences are important decision points for students to persist or leave their intended fields of study, it is essential to develop a predictive model to examine factors that impact STEM persistence in the context of chemistry. A robust longitudinal model would prove especially useful to assess stability of these factors, identify developmental trends and observe progressive changes (Ruspini, 1999).

Purpose of the study

The aim of this study was to develop and test a comprehensive persistence model that merges self-efficacy and outcome expectations with performance measures. As part of this, two questions were addressed:

1) What are the significant factors that predict students' performance while enrolled in general chemistry?
2) What are the significant factors that predict students' persistence in their intended STEM majors while enrolled in general chemistry?

Examining STEM persistence of students in the context of chemistry gateway courses will allow for the development of longitudinal models in not just chemistry but other physical sciences as well. This study will also utilize affective and performance measures to expand current knowledge and offer evidence regarding the participation and persistence differential in STEM. Moreover, examining STEM persistence as more than a dichotomous outcome is a much needed approach to obtained a richer and more comprehensive understanding of students' persistence in their STEM majors.

Research methodology

Research design

The performance and persistence models in this study were developed as 'proofs of concept' using a cross-sectional research design; while the original intent of this study was to develop longitudinal models, inadequacies in sample size limited the implementation of a longitudinal design and application of the relevant statistical method.

The performance model was developed by integrating measures of performance with measures of self-efficacy and outcome expectations (developed and validated in chapters 4 and 5). Using Toledo placement (TP) and standardized testing scores (ACT) for preliminary performance and pretesting of persistence measures (mean subscale scores from pre- CSEAS and COES), the predictive power of this performance model was tested. Performance was measured using students' final exam and course percentages. **Figure 8.1** shows the predictors and outcome variable used for developing the performance model.

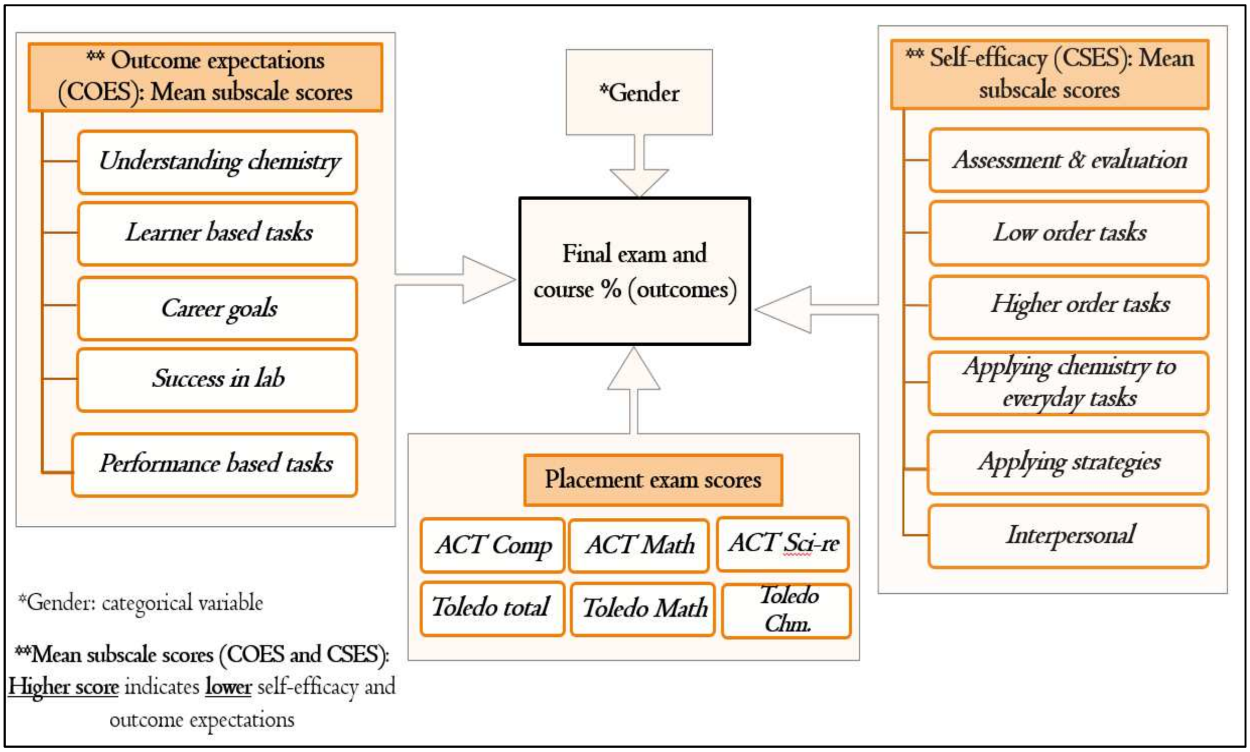


Figure 8.1. Outcome and predictors for developing and testing the performance model

The SCCT model of persistence was developed by examining students' long-term stability in a STEM major. While the predictors used in this model were identical to those used in the performance model, the outcome variable was categorical. Gender was used as a predictor in both models due to the historical relevance of gender in psychosocial models and career development. The models in this study utilized the following coding for gender: M (1) and F (2). **Figure 8.2** shows the predictors and outcome variable used for developing the persistence model.

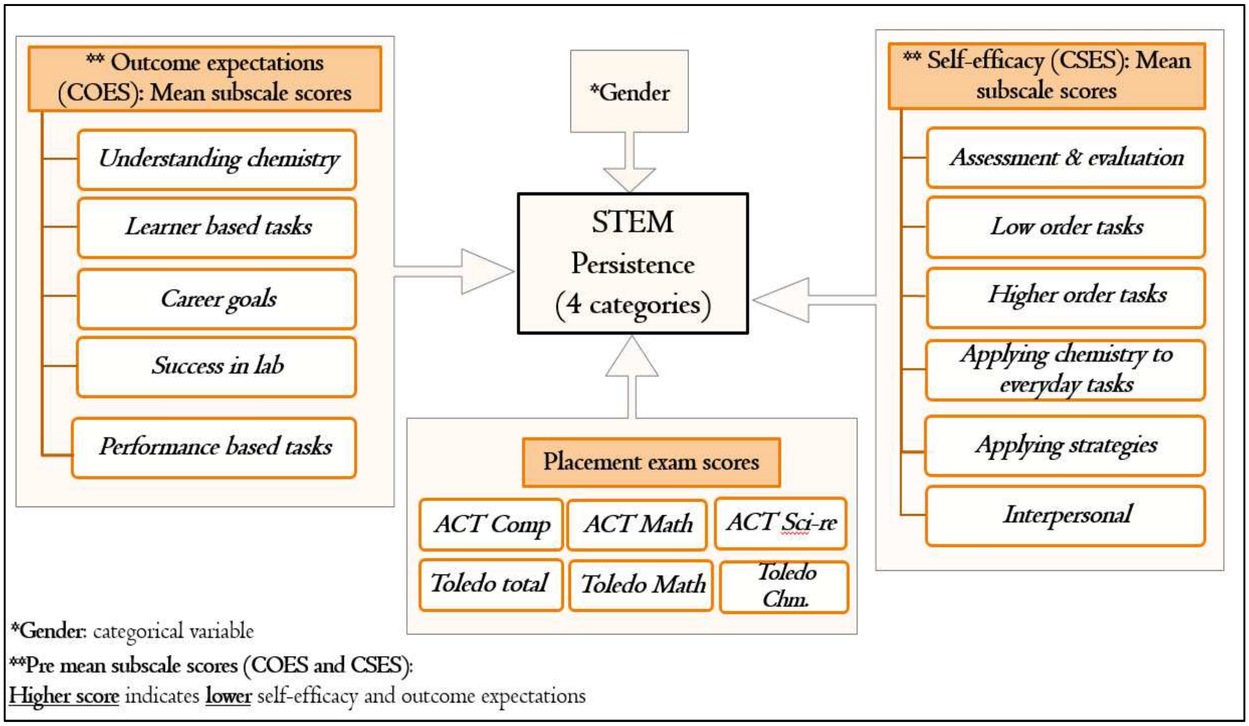


Figure 8.2. Outcome and predictors for developing and testing the persistence model

Before developing the persistence model, a few aspects regarding the outcome needed to be addressed. Students' majors at the start of GC I were coded as STEM (1), non-STEM (2) or Undecided (UND = 3). Different organizations and institutions offer varied lists indicating majors that can be placed in a STEM category. For the purposes of this study, the 2016 STEM Designated Degree Program List from the U.S. Immigration and Customs Enforcement (ICE) was used to assign majors as STEM or non-STEM. This list is available in **Appendix M.**

The outcome variable in the persistence model was STEM persistence, measured by the stability of a student's major. As the profile of students' goes beyond whether or not they stayed or did not stay in their intended STEM majors, this outcome was not a dichotomous variable. Instead, it was categorical and comprised of four groups that were coded based on whether a student's initial major from GC I (STEM, non-STEM or UND) changed or stayed in the same category at the end of GC I. The coding scheme to describe persistence is shown in the **Table 8.1.**

Table 8.1. Persistence categories and related codes

	Category	Code
a)	Student persisted in a STEM major	1
b)	Student switched into a STEM major (from undecided or non-STEM)	2
c)	Student stayed in a non-STEM major	3
d)	Student switched into a non-STEM major (from undecided or STEM major)	4

Participants

Each question posed in this study was addressed by using students enrolled in GC I at a large, urban, research intensive public university in the Midwestern United States.

Students considered in the originally intended longitudinal persistence and performance studies were those who would have started in GC I and ended in GC II the following semester. While the enrollment in GC II at any subsequent time point could have still constituted a longitudinal model, only those who took GC I and GC II in sequence would have been considered. In addition, students who enrolled in either course during summer sessions were excluded from this model.

Out of 608 IRB approved GC I students (Spring 2014 – Fall 2015), whose data had been cleaned based on the criteria described in chapter 3, 453 students enrolled in GC II the following semester. As persistence in STEM involved tracking students' STEM majors while enrolled in general chemistry, in order to be included in a longitudinal persistence model, the 608 students in GC I a) would have to be IRB approved, enrolled in GC II in sequence (the following semester) and stay through GC II, b) would have completed pre-affective measures in GCI, post-affective measures in GC II and pre-affective measures in GC II, c) indicated their intended major at the start of GC I and end of GC II, d) would have data available for their indicators of general cognitive ability (ACT Math, Sci-re, composite scores, TP Math and Chemistry scores, placement test scores) at the start of GCI and GC II, and e) indicators of performance (GC I final exam scores). When these criteria were applied, only 199 students could be used as part of the persistence model at the end of GC II. For the longitudinal performance study, only 130 students had all relevant variables and indices to be included in this study; several students had not taken the final exam and as a result could not be included in the performance model. These limitations resulted in developing these models using a cross-sectional design.

Students considered in the cross-sectional (referring to the same semester) persistence and performance studies were those who stayed enrolled in GC I during a semester (pre to post). Out of 608 IRB approved GC I students (Spring 2014 – Fall 2015), 523 students had usable data (pre-affective measures, ACT and TP scores at start of GCI) to develop the performance model based on the final exam score while 552 students had usable data for the performance model based on course percentages. These performance models were cross-validated using the data from Spring 2016 (103 students with final exam scores and 108 students with course percent data). For the performance models based on course performance, students who did not take the final exam were excluded from the regression analyses because course performance was heavily dependent on final exam scores. The persistence models were developed using data from Spring 2014 – Spring 2016 as the statistical method used for analyses mandated a large sample size. As a result, this model was developed but not tested for its predictive utility. There were 438 students with available data for the persistence model (pre-affective measures, ACT and TP scores at start of GCI, self-reported majors at start and end of GC I).

The discrepancy in the number of available students arises from the fact that while students took the final exam and obtained a score, they did not necessarily take the post survey administered two weeks before the final exam; this survey offered an opportunity to capture the students' most recent major and consequently if students did not complete this survey, there was no record of their major at the end of the tracking period either longitudinally or otherwise. Since these models had to be stable enough to make predictions or at least have enough cases to conduct cross-validation analysis, the cross-sectional design was used to develop each model.

Data analysis – Performance model

Standard multiple linear regression (SMLR) analysis was used to develop and test the performance models for the cross-sectional data set. The outcome variables were performance indicators – percentage on the final exam (ACS standardized exam) and in the course. This method was used to assess the size of the overall relationship between the performance indicators and predictor variables. In addition, the unique contribution of each predictor variable to the model was also assessed. Correlational analyses were also conducted to discover the significance of the predictor variables in contributing to the dependent variable. A correlation matrix served as a starting point to reveal significant associations between predictor and outcome variables. Partial correlations were also conducted to determine the effects of including confounding variables e.g. including both ACT Math and TP Math scores might not seem particularly beneficial since there could be shared variance between the two Math placement indicators. A way to confirm this would be to run a partial correlation of each predictor, while controlling for the co-variate, with the outcome variable and assess the relative impact of each predictor on the outcome.

Since there were several predictor variables, the starting point was to enter all possible predictor variables in the model. The second approach was to select variables that showed significant correlations with the outcome variable provided this correlation was stronger than the relationship between the selected predictors. Correlation coefficients, tolerance levels and the variance inflation factor (VIF) values between predictor variables were checked to ensure assumptions of multicollinearity had not been violated. In addition, emphasis was placed on normality of the residuals when assessing model fit. (Field, 2009).

Data analysis – Persistence model

Logistic regression, an example of a generalized linear model, allows for prediction of a discrete outcome such as category membership using predictor variables on any level of measurement. In logistic regression, the relationship between predictor and response variables is not a linear function but a logarithmic function (logit), in which 'probability' or 'odds' of the response assuming a particular value is assessed based on combination of values taken on by the predictor variables (Menard, 1995). While binary logistic regression is more prevalent and has dichotomous, probabilistic outcomes of 1 or 0, multinomial logistic regression, an extension of binary logistic regression, uses multiple independent variables to predict the probability of category membership in more than two categories of the dependent or outcome variable (Menard, 1995).

Although this method does not require fulfilment of normal distributions or linear relationships on the predictors in each group, it does require absence of multicollinearity; it is also assumed that category memberships are independent (Menard, 1995).

Since one of the research questions addressed in this chapter was how performance and affective measures affect persistence in a STEM major – with persistence having more than two levels – multinomial logistic regression was the best analytic approach to develop the persistence model for the cross-sectional data set. The standard logit model – with all predictors entered into the model at once – was selected in this regression analysis.

Although both models have several predictor variables, stepwise regression was not attempted due to its tendency to capitalize on chance and produce results that are often not generalizable to other similar samples (Field, 2009).

Descriptive statistics were obtained for relevant variables in both models. These analyses were performed using SPSS statistical software versions 23/24 and Excel 2015/2016.

Results and Discussion

Cross-sectional performance model – GC I final exam as outcome

Prior to examining the variables that impact longitudinal performance, descriptive statistics were obtained for all the variables in the model. **Table 8.2** shows descriptions, means, standard deviations and other statistics for each GC I pre- performance and affective measure that would potentially be included in the model. These statistics have been separated by gender to highlight any apparent differences based on the categorical variable in the model.

Table 8.2. Descriptive statistics (by sex) for variables in the performance model (data from S14 – F15)

	Variables	N	Missing	Mean	Std. deviaton	Min	Max	Skewness	Kurtosis
Males (1)	ACT Composite	220	60	23.66	3.52	14	34	.05	-.04
	ACT Math	220	60	23.61	3.90	14	34	-.17	-.18
	ACT Sci-Re	220	60	24.01	3.75	15	36	.42	.43
	TP - Math	273	7	84.60	10.52	50	100	-.65	-.05
	TP - Chemistry	273	7	65.92	10.31	25	93	-.40	.43
	TP - Total	273	7	72.13	8.81	42	93	-.28	-.16
	OE - Understanding	279	1	1.68	.45	1.00	3.33	.28	-.28
	OE - Performance based tasks	280	0	1.30	.41	1.00	2.67	1.24	.72
	OE - Career	280	0	1.63	.45	1.00	3.50	.53	.14
	OE - Lab tasks	280	0	1.75	.62	1.00	3.67	.65	-.18
	OE - Learner based tasks	280	0	2.85	.77	1.00	5.00	.15	.07
	SE - Assessment	280	0	1.90	.65	1.00	4.43	.87	1.24
	SE - Interpersonal	280	0	2.14	.82	1.00	5.00	.60	.28
	SE - Strategies and tasks	280	90	2.16	.68	1.00	5.00	1.04	1.59
	SE - Low order / recall tasks	280	90	2.13	.63	1.00	4.20	.54	.29
	SE - High order tasks	280	90	2.52	.78	1.00	5.00	.21	-.16
	SE - Apply chem. to everyday tasks	280	90	1.92	.72	1.00	5.00	.93	1.16
	GC I Final Exam %	242	38	72.17	14.74	32.50	100.00	-.40	-.49
	GC I Course %	255	25	82.50	16.89	1.90	113.38	-1.12	2.34
Females (2)	ACT Composite	266	62	23.36	3.46	15	34	.07	-.03
	ACT Math	266	62	22.70	4.09	14	34	.05	-.56
	ACT Sci-Re	266	62	23.17	3.42	11	35	-.05	1.11
	TP - Math	325	3	80.82	13.51	0	100	-1.56	5.42
	TP - Chemistry	325	3	60.21	10.52	23	90	-.15	.14
	TP - Total	325	3	67.08	9.78	28	92	-.50	.78
	OE - Understanding	328	0	1.73	.43	1.00	2.83	.00	-.89
	OE - Performance based tasks	328	0	1.38	.47	1.00	3.67	1.13	1.12
	OE - Career	328	0	1.57	.41	1.00	2.67	.39	-.61
	OE - Lab tasks	328	0	1.87	.73	1.00	4.00	.59	-.43
	OE - Learner based tasks	328	0	2.56	.73	1.00	5.00	.33	.02
	SE - Assessment	328	0	2.05	.72	1.00	4.57	.71	.50
	SE - Interpersonal	328	0	2.30	.80	1.00	4.67	.32	-.43
	SE - Strategies and tasks	328	0	2.30	.73	1.00	5.00	.74	.54
	SE - Low order / recall tasks	328	0	2.28	.72	1.00	4.80	.43	.00
	SE - High order tasks	327	1	2.68	.86	1.00	5.00	-.20	-.30
	SE - Apply chem. to everyday tasks	228	0	2.08	.79	1.00	5.00	.53	-.01
	GC I Final Exam %	281	47	66.53	14.50	11.25	99.50	-.35	-.02
	GC I Course %	297	31	79.58	16.50	3.30	107.84	-1.38	2.71

The descriptive statistics show some differences between means for certain variables in each group. While the significance of these differences is not displayed here, males and females showed significant differences in almost all affective measures except outcome expectations related to career and understanding chemistry. Among the performance indicators, there were significant differences between males and females in every placement test measure except ACT composite scores. Skewness and kurtosis values are at acceptable levels for most of the variables, although there are some variables in each group that exhibit considerable skewness and kurtosis. While these are criteria to consider when assessing predictors, there is no requirement that variables be normally distributed in multiple regression. The more important distributional assumption is for model errors, so the analyses were carried out without any transformations to these variables.

One of the first steps to determine which variables had to be included in the model was to conduct a correlation analysis. A predictor was considered inclusionary if it exhibited a significant correlation with the outcome variable. The correlation matrix (for this model) displaying significant bivariate correlations among predictor variables and the GC I final exam score is shown in **Table 8.3**. The mean and standard deviation of each variable is also indicated.

Among the performance indicators, the strongest and significant correlations to GC I final exam score resulted from the placement measures: ACT Composite (r= .469**), ACT Math (r = .440**), ACT Sci-Re (r=.411**), TP Math (r= .475**), TP Chem. (r=.498**) and TP total (r=.575**). Among the persistence measures, there were significant correlations between OE – learner based tasks (r= -.127**), OE – understanding (r= -.097*), SE – exam preparation (r= -.206**), SE – general strategies and tasks (r= -.215**), SE – low order tasks (r= -.154**) and SE – applying chemistry to everyday tasks (r= -.249**) and the final exam score respectively.

Table 8.3. Bivariate correlations between predictor variables and GC I final exam % for cross-sectional performance model. Higher mean SE and OE subscale scores indicate low SE and OE respectively.

		GC I Final Exam %	ACT Comp	ACT Math	ACT Sci-Re	TP - Math	TP - Chem	TP - total	OE - Understanding	OE - Performance based tasks	OE - Career	OE - Lab	OE - Learner based	SE - Exam prep	SE - Interpersonal	SE - general strategies and tasks	SE - Low order tasks	SE - Higher order tasks	SE_apply chemistry to everyday tasks	Mean	Std. deviation
GC I Final Exam %	Pearson Correlation	1	.469**	.440**	.411**	.475**	.498**	.575**	-.097*	-0.093	0.038	-0.092	-.127**	-.206**	-0.080	-.215**	-.154**	-0.073	-.249**	68.64	14.16
	Sig. (2-tailed)		0.000	0.000	0.000	0.000	0.000	0.000	0.048	0.058	0.439	0.059	0.010	0.000	0.101	0.000	0.002	0.134	0.000		
ACT Comp	Pearson Correlation	.469**	1	.798**	.826**	.369**	.315**	.392**	0.009	-0.034	0.056	-0.043	-.195**	-0.096	-0.035	-.117*	0.030	-0.037	-.132**	23.55	3.53
	Sig. (2-tailed)	0.000		0.000	0.000	0.000	0.000	0.000	0.850	0.486	0.252	0.386	0.000	0.050	0.474	0.017	0.543	0.452	0.007		
ACT Math	Pearson Correlation	.440**	.798**	1	.655**	.437**	.340**	.440**	0.005	0.004	0.052	-0.055	-0.096	-.123*	-0.026	-.104*	0.034	-0.008	-.165**	23.17	4.03
	Sig. (2-tailed)	0.000	0.000		0.000	0.000	0.000	0.000	0.922	0.943	0.287	0.263	0.051	0.012	0.601	0.034	0.484	0.877	0.001		
ACT Sci-Re	Pearson Correlation	.411**	.826**	.655**	1	.345**	.291**	.364**	-0.007	-0.072	0.064	-0.086	-.118*	-0.073	-0.013	-0.064	0.078	-0.004	-0.082	23.66	3.65
	Sig. (2-tailed)	0.000	0.000	0.000		0.000	0.000	0.000	0.879	0.144	0.195	0.079	0.016	0.135	0.795	0.193	0.111	0.932	0.093		
TP - Math	Pearson Correlation	.475**	.369**	.437**	.345**	1	.398**	.717**	-0.059	-0.071	-0.016	-0.012	-0.032	-.144**	-0.021	-.110*	0.008	0.000	-.186**	83.46	11.49
	Sig. (2-tailed)	0.000	0.000	0.000	0.000		0.000	0.000	0.230	0.146	0.749	0.802	0.509	0.003	0.667	0.025	0.866	0.994	0.000		
TP - Chem	Pearson Correlation	.498**	.315**	.340**	.291**	.398**	1	.925**	-0.037	-0.041	0.032	-0.017	-0.057	-.163**	-0.028	-.198**	-.225**	-.131**	-.289**	63.78	10.56
	Sig. (2-tailed)	0.000	0.000	0.000	0.000	0.000		0.000	0.456	0.406	0.511	0.727	0.244	0.001	0.568	0.000	0.000	0.007	0.000		
TP - total	Pearson Correlation	.575**	.392**	.440**	.364**	.717**	.925**	1	-0.052	-0.061	0.018	-0.018	-0.058	-.183**	-0.030	-.197**	-.167**	-.100*	-.297**	70.33	9.25
	Sig. (2-tailed)	0.000	0.000	0.000	0.000	0.000	0.000		0.285	0.211	0.719	0.706	0.238	0.000	0.542	0.000	0.001	0.042	0.000		
OE - Understanding	Pearson Correlation	-.097*	0.009	0.005	-0.007	-0.059	-0.037	-0.052	1	.569**	.507**	.514**	-0.092	.411**	.294**	.164**	.231**	.185**	.193**	1.71	0.44
	Sig. (2-tailed)	0.048	0.850	0.922	0.879	0.230	0.456	0.285		0.000	0.000	0.000	0.060	0.000	0.000	0.001	0.000	0.000	0.000		
OE - Performance based tasks	Pearson Correlation	-0.093	-0.034	0.004	-0.072	-0.071	-0.041	-0.061	.569**	1	.564**	.434**	-0.012	.392**	.310**	.241**	.246**	.135**	.175**	1.34	0.44
	Sig. (2-tailed)	0.058	0.486	0.943	0.144	0.146	0.406	0.211	0.000		0.000	0.000	0.805	0.000	0.000	0.000	0.000	0.006	0.000		
OE - Career	Pearson Correlation	0.038	0.056	0.052	0.064	-0.016	0.032	0.018	.507**	.564**	1	.363**	-0.023	.303**	.270**	.111*	.106*	.146**	.160**	1.60	0.43
	Sig. (2-tailed)	0.439	0.252	0.287	0.195	0.749	0.511	0.719	0.000	0.000		0.000	0.645	0.000	0.000	0.024	0.031	0.003	0.001		
OE - Lab	Pearson Correlation	-0.092	-0.043	-0.055	-0.086	-0.012	-0.017	-0.018	.514**	.434**	.363**	1	-.128**	.299**	.262**	.100*	.196**	.142**	.129**	1.84	0.68
	Sig. (2-tailed)	0.059	0.386	0.263	0.079	0.802	0.727	0.706	0.000	0.000	0.000		0.009	0.000	0.000	0.042	0.000	0.004	0.009		
OE - Learner based	Pearson Correlation	-.127**	-.195**	-0.096	-.118*	-0.032	-0.057	-0.058	-0.092	-0.012	-0.023	-.128**	1	-0.023	-0.011	0.011	-0.074	-.098*	-0.033	3.27	0.74
	Sig. (2-tailed)	0.010	0.000	0.051	0.016	0.509	0.244	0.238	0.060	0.805	0.645	0.009		0.640	0.816	0.820	0.133	0.045	0.500		
SE - Exam prep	Pearson Correlation	-.206**	-0.096	-.123*	-0.073	-.144**	-.163**	-.183**	.411**	.392**	.303**	.299**	-0.023	1	.592**	.484**	.427**	.361**	.350**	4.03	0.69
	Sig. (2-tailed)	0.000	0.050	0.012	0.135	0.003	0.001	0.000	0.000	0.000	0.000	0.000	0.640		0.000	0.000	0.000	0.000	0.000		
SE - Interpersonal	Pearson Correlation	-0.080	-0.035	-0.026	-0.013	-0.021	-0.028	-0.030	.294**	.310**	.270**	.262**	-0.011	.592**	1	.237**	.296**	.318**	.220**	3.74	0.83
	Sig. (2-tailed)	0.101	0.474	0.601	0.795	0.667	0.568	0.542	0.000	0.000	0.000	0.000	0.816	0.000		0.000	0.000	0.000	0.000		
SE - general strategies and tasks	Pearson Correlation	-.215**	-.117*	-.104*	-0.064	-.110*	-.198**	-.197**	.164**	.241**	.111*	.100*	0.011	.484**	.237**	1	.492**	.243**	.415**	3.78	0.69
	Sig. (2-tailed)	0.000	0.017	0.034	0.193	0.025	0.000	0.000	0.001	0.000	0.024	0.042	0.820	0.000	0.000		0.000	0.000	0.000		
SE - Low order tasks	Pearson Correlation	-.154**	0.030	0.034	0.078	0.008	-.225**	-.167**	.231**	.246**	.106*	.196**	-0.074	.427**	.296**	.492**	1	.477**	.434**	3.82	0.65
	Sig. (2-tailed)	0.002	0.543	0.484	0.111	0.866	0.000	0.001	0.000	0.000	0.031	0.000	0.133	0.000	0.000	0.000		0.000	0.000		
SE - Higher order tasks	Pearson Correlation	-0.073	-0.037	-0.008	-0.004	0.000	-.131**	-.100*	.185**	.135**	.146**	.142**	-.098*	.361**	.318**	.243**	.477**	1	.402**	3.39	0.83
	Sig. (2-tailed)	0.134	0.452	0.877	0.932	0.994	0.007	0.042	0.000	0.006	0.003	0.004	0.045	0.000	0.000	0.000	0.000		0.000		
SE_apply chemistry to everyday tasks	Pearson Correlation	-.249**	-.132**	-.165**	-0.082	-.186**	-.289**	-.297**	.193**	.175**	.160**	.129**	-0.033	.350**	.220**	.415**	.434**	.402**	1	3.99	0.75
	Sig. (2-tailed)	0.000	0.007	0.001	0.093	0.000	0.000	0.000	0.000	0.000	0.001	0.009	0.500	0.000	0.000	0.000	0.000	0.000			

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).
Listwise N=418

However, examining the matrix closely indicated that while the performance indicators had strong, significant correlations with the outcome variable, some of the inter-correlations among the predictors themselves were significantly higher. These indications of potential multicollinearity were confirmed when all the predictors were entered into the multiple regression model and the VIF values for TP Math, TP Chemistry and TP total were 973.70, 3274.45 and 5667.21 respectively. When TP total was excluded from the model, VIF values returned to acceptable levels. While the ACT variables did not display values as high as the TP variables, the VIF and tolerance values were still violating assumptions of multicollinearity, with ACT Composite making the highest contribution to these violations. Thus, while both ACT Composite and TP total could be included together in a model, neither could be included along with their individual subscores. Normal P-P plots for all predictors indicated a reasonably straight line.

ACT composite scores and total TP scores had the strongest correlations with the final exam score (.469 and .575 respectively), while displaying a moderate correlation between themselves (.392). Among the affective measures, the self-efficacy subscales showed moderate correlations with the final exam score, with SE-assessment and evaluation and SE-applying strategies showing strong correlations with the outcome than with the placement test scores. These predictors were used as the starting points for developing the multiple regression model.

In addition to these evaluations, partial correlations were also examined to assess the relative impact of each predictor. Using the correlation matrix, in conjunction with partial correlations, resulted in a model three predictors that accounted for 40 % of the variance in the model as shown in **Table 8.4**.

Table 8.4. Summary of multiple regression analysis for students' final exam scores in GC I (N = 420).

Variables	B	SE (B)	β	t	Sig. (*p*)	Zero-order	Partial	Part	Tolerance	VIF
TP - total	.682	.064	.445	10.742	.000	.575	.466	.404	.824	1.214
ACT - Comp.	1.146	.164	.286	6.973	.000	.470	.324	.263	.845	1.184
SE - exam prep	-1.980	.784	-.097	-2.525	.012	-.204	-.123	-.095	.967	1.035
R^2 = .410 ; Adj. R^2 = .406 F(3,419) = 96.523, p < .001 Average of residuals = .003	GC I final exam % = - 2.42 + (.628 TP total) + (1.146 ACT-Comp) + (-1.980 SE-exam prep)									

The model was statistically significant, $F(3, 419) = 96.523$, $p < .001$. The unstandardized coefficients (B) provide information about the relationship between the final exam score and each predictor. In this model, as total TP scores increase by one point, the final exam score increases by 0.682 points; as ACT composite scores increase by one point, the final exam score increases by 1.146 points and since in the self-efficacy scale, a higher mean subscale score indicates lower self-efficacy, a lower self-efficacy related to exam preparation and assessment decreases the final exam score by 1.98 points. While the model was fair, as indicated by its R^2 value, it should be noted that the affective measure captures perceived self-efficacy at the start of the course, and is not very closely tied to the outcome. Given this situation, it is expected that past performance indicators (ability) would be the strongest contributors to the performance model.

Standardized regression residual plots, as displayed in **Figure 8.3**, showed most of the residual values around zero with no obvious 'funneling', thus homoscedasticity was assumed. In addition, the average of residuals was .003, normal P-P plot of regression standardized residual did not show deviations from the straight line and normality tests conducted on the residuals were not significant, indicating that residuals were normal. There were no apparent outliers observed in the residual plots or influence statistics. Although there was one case that exceeded the critical value of 18.47 for Mahalanobis distances, removal of this case did not alter the regression model.

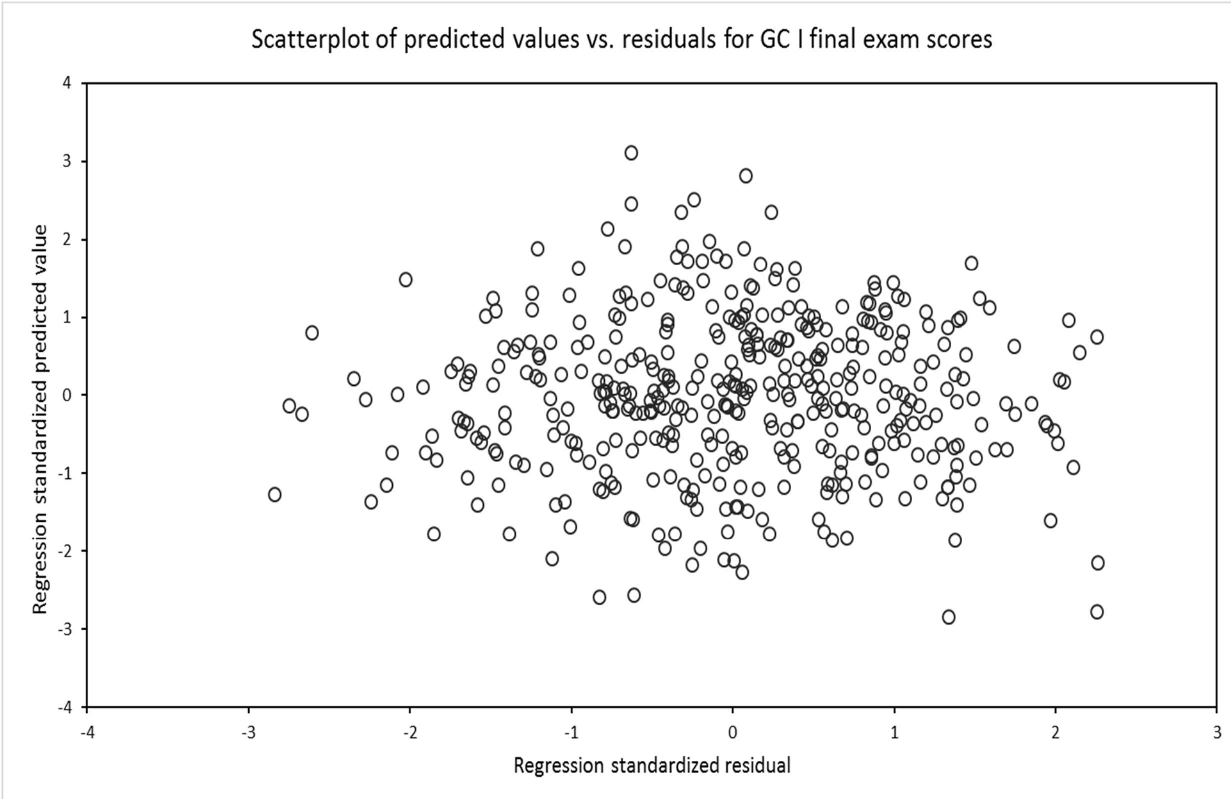


Figure 8.3. Scatter plot distribution of residuals for performance model using GC I final exam scores

Since the model displayed a fairly moderate R-squared value, a cross-validation analysis was conducted using this model on a data set of 89 students (with available data) from Spring 2016. However, when this process was implemented, the average of the residuals calculated from this data set was -2.373, with an R-squared value of 50.7%, indicating an over-estimation of the model. The small sample size could have contributed to this since sample size and ratio of predictors to sample size can over-estimate or shrink the values of regression predictors, resulting in biased R^2 values. (Copas, 1987). It is also possible that the tests in the spring term were more difficult or students in the spring term were less motivated to succeed in te course than students in the fall term. Although not implemented here, alternate procedures such as bootstrapping or jackknifing would have provided better estimates for R^2 for a test data set (Browne, 2000).

Cross-sectional performance model – GC I course performance as outcome

Using the same process as before, the first step was to obtain a correlation matrix displaying correlations between predictors and the outcome variable – GC I course performance. This matrix is shown in **Table 8.5**.

Table 8.5. Bivariate correlations between predictor variables and GC I course % for cross-sectional performance model. Higher mean SE and OE subscale scores indicate low SE and OE respectively.

		GC I Course %	ACT Comp	ACT Math	ACT Sci-Re	TP - Math	TP - Chem	TP - total	OE - Understanding	OE - Performance based tasks	OE - Career	OE - Lab	OE - Learner based	SE - Exam prep	SE - Interpersonal	SE - general strategies and tasks	SE - Low order tasks	SE - Higher order tasks	SE_apply chemistry to everyday tasks	Mean	Std. deviation
GC I Course %	Pearson Correlation	1	.254**	.234**	.243**	.265**	.293**	.331**	-.108*	-.110*	-0.018	-.109*	-0.088	-.188**	-0.085	-.154**	-.137**	-0.055	-.149**	80.78	15.59
	Sig. (2-tailed)		0.000	0.000	0.000	0.000	0.000	0.000	0.023	0.021	0.701	0.023	0.065	0.000	0.074	0.001	0.004	0.253	0.002		
ACT Comp	Pearson Correlation	.254**	1	.800**	.828**	.382**	.324**	.404**	-0.004	-0.067	0.051	-0.032	-.194**	-.111*	-0.035	-.126**	0.022	-0.045	-.134**	23.54	3.52
	Sig. (2-tailed)	0.000		0.000	0.000	0.000	0.000	0.000	0.926	0.158	0.288	0.499	0.000	0.020	0.463	0.008	0.642	0.344	0.005		
ACT Math	Pearson Correlation	.234**	.800**	1	.662**	.448**	.353**	.454**	-0.006	-0.024	0.055	-0.044	-.101*	-.130**	-0.031	-.111*	0.031	-0.013	-.165**	23.18	4.02
	Sig. (2-tailed)	0.000	0.000		0.000	0.000	0.000	0.000	0.903	0.618	0.251	0.354	0.034	0.006	0.511	0.019	0.516	0.785	0.000		
ACT Sci-Re	Pearson Correlation	.243**	.828**	.662**	1	.353**	.298**	.373**	-0.024	-.112*	0.047	-0.083	-.120*	-.098*	-0.023	-0.074	0.063	-0.022	-0.082	23.64	3.64
	Sig. (2-tailed)	0.000	0.000	0.000		0.000	0.000	0.000	0.616	0.019	0.321	0.080	0.012	0.041	0.633	0.122	0.188	0.645	0.087		
TP - Math	Pearson Correlation	.265**	.382**	.448**	.353**	1	.401**	.720**	-0.064	-0.087	-0.014	-0.005	-0.036	-.146**	-0.022	-.115*	0.018	0.013	-.186**	83.50	11.58
	Sig. (2-tailed)	0.000	0.000	0.000	0.000		0.000	0.000	0.183	0.068	0.775	0.913	0.450	0.002	0.648	0.016	0.701	0.788	0.000		
TP - Chem	Pearson Correlation	.293**	.324**	.353**	.298**	.401**	1	.924**	-0.032	-0.036	0.045	0.003	-0.047	-.165**	-0.028	-.214**	-.236**	-.139**	-.299**	63.75	10.55
	Sig. (2-tailed)	0.000	0.000	0.000	0.000	0.000		0.000	0.500	0.455	0.351	0.953	0.329	0.001	0.561	0.000	0.000	0.003	0.000		
TP - total	Pearson Correlation	.331**	.404**	.454**	.373**	.720**	.924**	1	-0.051	-0.064	0.028	0.000	-0.051	-.186**	-0.030	-.210**	-.171**	-.100*	-.305**	70.35	9.27
	Sig. (2-tailed)	0.000	0.000	0.000	0.000	0.000	0.000		0.283	0.180	0.562	0.993	0.283	0.000	0.530	0.000	0.000	0.036	0.000		
OE - Understanding	Pearson Correlation	-.108*	-0.004	-0.006	-0.024	-0.064	-0.032	-0.051	1	.556**	.492**	.516**	-0.089	.417**	.293**	.151**	.215**	.167**	.167**	1.71	0.44
	Sig. (2-tailed)	0.023	0.926	0.903	0.616	0.183	0.500	0.283		0.000	0.000	0.000	0.064	0.000	0.000	0.001	0.000	0.000	0.000		
OE - Performance based tasks	Pearson Correlation	-.110*	-0.067	-0.024	-.112*	-0.087	-0.036	-0.064	.556**	1	.573**	.418**	0.007	.414**	.302**	.230**	.244**	.138**	.156**	1.35	0.45
	Sig. (2-tailed)	0.021	0.158	0.618	0.019	0.068	0.455	0.180	0.000		0.000	0.000	0.880	0.000	0.000	0.000	0.000	0.004	0.001		
OE - Career	Pearson Correlation	-0.018	0.051	0.055	0.047	-0.014	0.045	0.028	.492**	.573**	1	.370**	-0.021	.307**	.273**	.115*	.108*	.149**	.133**	1.60	0.44
	Sig. (2-tailed)	0.701	0.288	0.251	0.321	0.775	0.351	0.562	0.000	0.000		0.000	0.656	0.000	0.000	0.016	0.023	0.002	0.005		
OE - Lab	Pearson Correlation	-.109*	-0.032	-0.044	-0.083	-0.005	0.003	0.000	.516**	.418**	.370**	1	-.124**	.285**	.266**	0.084	.171**	.122*	.105*	1.83	0.68
	Sig. (2-tailed)	0.023	0.499	0.354	0.080	0.913	0.953	0.993	0.000	0.000	0.000		0.009	0.000	0.000	0.077	0.000	0.010	0.027		
OE - Learner based	Pearson Correlation	-0.088	-.194**	-.101*	-.120*	-0.036	-0.047	-0.051	-0.089	0.007	-0.021	-.124**	1	-0.025	-0.018	0.003	-0.082	-.103*	-0.038	2.72	0.74
	Sig. (2-tailed)	0.065	0.000	0.034	0.012	0.450	0.329	0.283	0.064	0.880	0.656	0.009		0.600	0.705	0.951	0.086	0.030	0.421		
SE - Exam prep	Pearson Correlation	-.188**	-.111*	-.130**	-.098*	-.146**	-.165**	-.186**	.417**	.414**	.307**	.285**	-0.025	1	.580**	.478**	.428**	.360**	.340**	1.97	0.69
	Sig. (2-tailed)	0.000	0.020	0.006	0.041	0.002	0.001	0.000	0.000	0.000	0.000	0.000	0.600		0.000	0.000	0.000	0.000	0.000		
SE - Interpersonal	Pearson Correlation	-0.085	-0.035	-0.031	-0.023	-0.022	-0.028	-0.030	.293**	.302**	.273**	.266**	-0.018	.580**	1	.235**	.288**	.320**	.209**	2.26	0.82
	Sig. (2-tailed)	0.074	0.463	0.511	0.633	0.648	0.561	0.530	0.000	0.000	0.000	0.000	0.705	0.000		0.000	0.000	0.000	0.000		
SE - general strategies and tasks	Pearson Correlation	-.154**	-.126**	-.111*	-0.074	-.115*	-.214**	-.210**	.151**	.230**	.115*	0.084	0.003	.478**	.235**	1	.491**	.258**	.421**	2.21	0.69
	Sig. (2-tailed)	0.001	0.008	0.019	0.122	0.016	0.000	0.000	0.001	0.000	0.016	0.077	0.951	0.000	0.000		0.000	0.000	0.000		
SE - Low order tasks	Pearson Correlation	-.137**	0.022	0.031	0.063	0.018	-.236**	-.171**	.215**	.244**	.108*	.171**	-0.082	.428**	.288**	.491**	1	.495**	.437**	2.18	0.65
	Sig. (2-tailed)	0.004	0.642	0.516	0.188	0.701	0.000	0.000	0.000	0.000	0.023	0.000	0.086	0.000	0.000	0.000		0.000	0.000		
SE - Higher order tasks	Pearson Correlation	-0.055	-0.045	-0.013	-0.022	0.013	-.139**	-.100*	.167**	.138**	.149**	.122*	-.103*	.360**	.320**	.258**	.495**	1	.398**	2.61	0.83
	Sig. (2-tailed)	0.253	0.344	0.785	0.645	0.788	0.003	0.036	0.000	0.004	0.002	0.010	0.030	0.000	0.000	0.000	0.000		0.000		
SE_apply chemistry to everyday tasks	Pearson Correlation	-.149**	-.134**	-.165**	-0.082	-.186**	-.299**	-.305**	.167**	.156**	.133**	.105*	-0.038	.340**	.209**	.421**	.437**	.398**	1	2.00	0.75
	Sig. (2-tailed)	0.002	0.005	0.000	0.087	0.000	0.000	0.000	0.000	0.001	0.005	0.027	0.421	0.000	0.000	0.000	0.000	0.000			

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).
Listwise N=440

As the outcome was different for this model and used a different sample of students (excluding those who had not taken the final exam), the correlation matrix had to be replicated. Among the performance indicators, the moderately significant correlations to GC I course performance resulted from the placement measures: ACT Composite (r= .254**), ACT Math (r = .234**), ACT Sci-Re (r=.243**), TP Math (r= .265**), TP Chem. (r=.293**) and TP total (r=.331**). Among the persistence measures, there were significant correlations between OE – performance based tasks (r= -.110*), OE – understanding (r= -.108*), OE – lab (r=-.109*), SE – exam preparation (r= -.188**), SE – general strategies and tasks (r= -.154**), SE – low order tasks (r= -.137**) and SE – applying chemistry to everyday tasks (r= -.149**) and the course performance respectively.

However, similar to the final exam performance model, examining the matrix closely indicated potential multicollinearity among some of the performance indicators measuring general cognitive ability. These indications were confirmed when all the predictors were entered into the multiple regression model and the VIF values for TP Math, TP Chemistry and TP total were 1022.342, 3374.50 and 5885.19 respectively. When TP total was excluded entirely or included in the model by itself, VIF values for remaining variables returned to acceptable levels. In addition, the inter-correlations among placement test indices were significantly higher than their respective correlations with the outcome variable. Consequently, when selecting a placement test predictor for the model, the variable that showed the highest correlation with GC I course performance was selected to avoid problems with multicollinearity. Thus, in this model, the only placement test indicator included was the TP total score.

Using the TP total score, correlation matrix and partial correlations as a starting point, the model developed consisted of four predictors that accounted for a variance of 26% in the model as shown in **Table 8.6**.

Table 8.6. Summary of multiple regression analysis for students' course performance in GC I (N = 510).

Variables	B	SE (B)	β	t	Sig. (*p*)	Zero-order	Partial	Part	Tolerance	VIF
TP - total	.566	.053	.418	10.578	.000	.462	.426	.404	.934	1.071
OE - lab	-2.583	.706	-.142	-3.659	.000	-.155	-.161	-.140	.967	1.034
OE - learner based tasks	-2.422	.642	-.146	-3.773	.000	-.151	-.166	-.144	.978	1.023
SE - strategies and tasks	-2.051	.704	-.116	-2.913	.004	-.234	-.129	-.111	.925	1.081

R^2 = .264 ; Adj. R^2 = .258
$F(4,509)$ = 45.284, $p < .001$
Average of residuals = -.02

GC I course % = 59.784 + (.566 TP total) + (-2.583 OE-lab) + (-2.422 OE-learner based tasks) + (-2.051 SE-strategies)

The model was statistically significant, $F(4, 509) = 45.284$, $p < .001$. The unstandardized coefficients (B) provide information about the relationship between the final exam score and each predictor. In this model, as total TP scores increase by one point, the course percentage increases by 0.566 points. In the self-efficacy and outcome expectations scales, a higher mean subscale score indicates lower self-efficacy and less positive outcome expectations. Thus, in this model a lower self-efficacy related to applying general problem solving strategies and tasks decreases the course percentage by 2.05 points while lower expectations about lab and learner based tasks decrease the course percentage by 2.583 and 2.422 points respectively.

Since the model was now concerned with predicting course performance, it makes substantive sense that components such as general chemistry lab that contribute to the overall grade in the course become significant contributors to the model. In addition, students' expectations related to learner based tasks such as memorizing information and formulas and their confidence in being able to use problem solving strategies are important predictors of their course performance. Given that these are pre persistence measures and that even without the pre performance measure (TP total score), this model still accounts for ~ 10% variance is an indication that regardless of past performance or ability, there are affective factors at play when examining students' accomplishments and persistence.

Standardized regression residual plots, as displayed in **Figure 8.4**, showed most of the residual values around zero with no obvious 'funneling', thus homoscedasticity was assumed. In addition, the average of residuals was -.02, normal P-P plot of regression standardized residual showed slight deviations from the straight line and normality tests conducted on the residuals were not significant, indicating that residuals were normal. There were no apparent outliers observed in the residual plots or influence statistics.

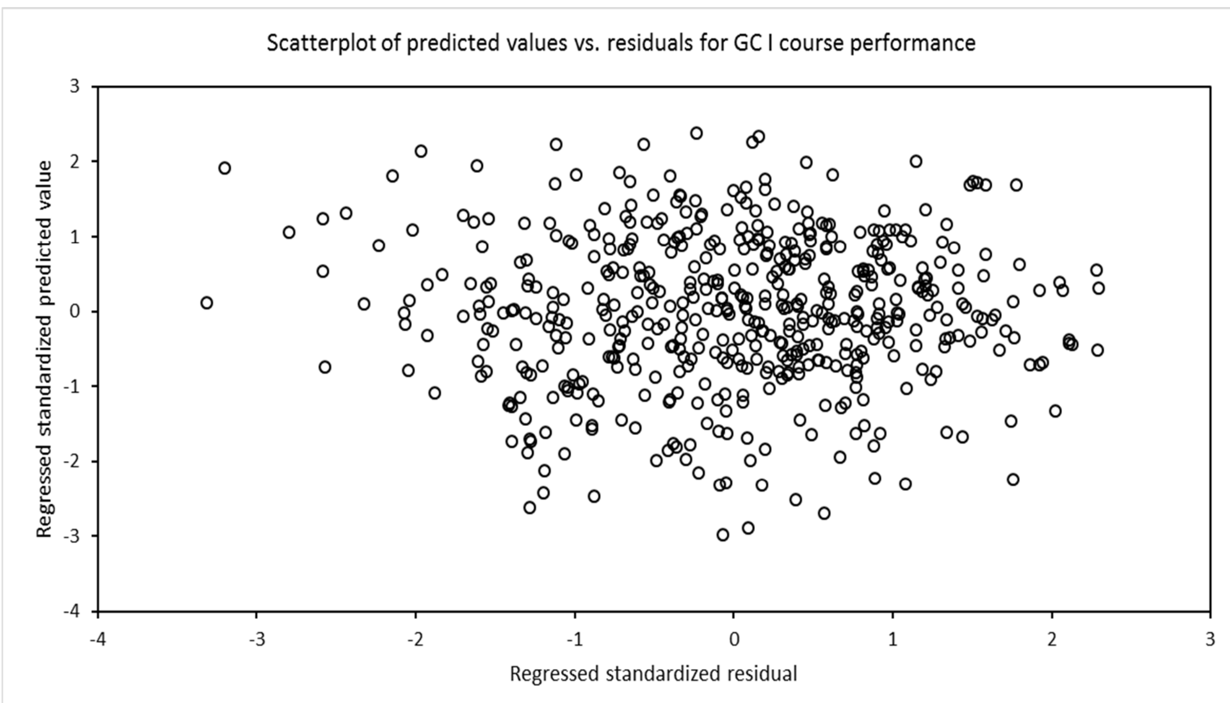


Figure 8.4. Scatter plot distribution of residuals for performance model using GC I course percentages

Cross validation using this model on the Spring 2016 data set resulted in the same problems of overestimation as with the performance model using the final exam. Thus, larger sample sizes or alternate techniques would offer better estimates for R^2.

Cross-sectional persistence model

Before conducting the analyses, descriptive statistics were obtained for pre- performance and persistence measures for students in each category. These statistics are shown in **Table 8.7a** and **Table 8.7b**.

Table 8.7a. Descriptive statistics (by category) for variables in the persistence model (Spring 2014 – Spring 2016)

	Variables	N	Missing	Mean	Std. deviaton	Min	Max	Skewness	Kurtosis
Category 1 - Stayed in STEM	Gender (M=1 and F=2)	315	70	1.56	0.50	1	2	-0.23	-1.96
	ACT Composite	319	66	23.45	3.53	14	33	-0.06	-0.24
	ACT Math	319	66	22.97	3.98	14	34	-0.04	-0.59
	ACT Sci-Re	319	66	23.46	3.66	11	36	0.07	0.74
	TP - Math	379	6	82.20	12.60	0	100	-1.38	4.70
	TP - Chemistry	379	6	62.92	10.43	33	90	-0.26	-0.11
	TP - Total	379	6	69.34	9.39	37	90	-0.51	0.32
	OE - Understanding	384	1	1.71	0.44	1.00	2.83	0.12	-0.77
	OE - Performance based tasks	385	0	1.35	0.45	1.00	2.67	1.08	0.28
	OE - Career	385	0	1.58	0.44	1.00	3.67	0.75	1.20
	OE - Lab tasks	385	0	1.82	0.67	1.00	4.00	0.67	-0.08
	OE - Learner based tasks	385	0	2.74	0.77	1.00	5.00	0.10	-0.45
	SE - Assessment	385	0	2.05	0.72	1.00	4.43	0.66	0.33
	SE - Interpersonal	385	0	2.31	0.82	1.00	5.00	0.32	-0.32
	SE - Strategies and tasks	385	0	2.23	0.65	1.00	4.33	0.83	0.80
	SE - Low order / recall tasks	385	0	2.21	0.68	1.00	4.80	0.60	0.31
	SE - High order tasks	384	1	2.68	0.84	1.00	5.00	0.23	-0.28
	SE - Apply chem. to everyday tasks	385	0	2.05	0.77	1.00	5.00	0.62	0.07
Category 2 - Switched into STEM	Gender (M=1 and F=2)	10	2	1.70	0.48	1	2	-1.04	-1.22
	ACT Composite	10	2	24.90	5.51	19	34	0.54	-1.18
	ACT Math	10	2	24.20	4.08	18	30	-0.07	-1.53
	ACT Sci-Re	10	2	25.70	4.30	19	35	0.85	1.96
	TP - Math	12	0	88.33	9.13	70	100	-0.70	0.19
	TP - Chemistry	12	0	69.58	10.60	53	90	0.12	0.08
	TP - Total	12	0	75.83	8.49	60	92	-0.05	0.49
	OE - Understanding	12	0	1.54	0.30	1.00	2.00	-0.01	-0.21
	OE - Performance based tasks	12	0	1.22	0.41	1.00	2.33	2.17	4.77
	OE - Career	12	0	1.51	0.34	1.00	2.00	-0.05	-1.61
	OE - Lab tasks	12	0	1.61	0.58	1.00	2.67	0.43	-1.10
	OE - Learner based tasks	12	0	2.72	0.91	1.33	4.33	0.02	-0.72
	SE - Assessment	12	0	1.75	0.53	1.00	2.86	0.94	0.39
	SE - Interpersonal	12	0	2.55	0.78	1.33	3.67	-0.23	-1.50
	SE - Strategies and tasks	12	0	1.92	0.78	1.00	4.00	1.79	4.46
	SE - Low order / recall tasks	12	0	1.98	0.62	1.40	3.80	2.60	7.96
	SE - High order tasks	12	0	2.39	0.90	1.00	4.33	0.36	1.34
	SE - Apply chem. to everyday tasks	12	0	1.86	0.87	1.00	4.00	1.25	2.37

Table 8.7b (Continued). Descriptive statistics (by category) for variables in the persistence model (Spring 2014 – Spring 2016)

	Variables	N	Missing	Mean	Std. deviaton	Min	Max	Skewness	Kurtosis
Category 3 - Stayed in non STEM	Gender (M=1 and F=2)	25	6	1.48	0.51	1	2	0.09	-2.17
	ACT Composite	24	7	23.00	3.28	17	29	-0.30	-0.52
	ACT Math	24	7	22.54	3.36	16	29	-0.32	-0.37
	ACT Sci-Re	24	7	23.17	3.07	18	29	-0.17	-0.51
	TP - Math	30	1	82.83	11.72	40	100	-1.69	5.13
	TP - Chemistry	30	1	59.75	12.34	23	85	-0.71	2.08
	TP - Total	30	1	67.44	11.09	28	90	-1.26	4.74
	OE - Understanding	31	0	1.76	0.44	1.00	2.50	-0.09	-0.89
	OE - Performance based tasks	31	0	1.32	0.50	1.00	2.67	1.53	1.59
	OE - Career	31	0	1.61	0.46	1.00	2.67	0.27	-0.75
	OE - Lab tasks	31	0	1.92	0.72	1.00	3.67	0.73	0.11
	OE - Learner based tasks	31	0	2.81	0.80	1.33	5.00	0.87	1.31
	SE - Assessment	31	0	2.00	0.84	1.00	4.57	1.06	1.48
	SE - Interpersonal	31	0	2.33	0.95	1.00	4.33	0.33	-0.63
	SE - Strategies and tasks	31	0	2.54	0.85	1.00	5.00	0.66	0.98
	SE - Low order / recall tasks	31	0	2.40	0.68	1.00	4.40	0.41	1.57
	SE - High order tasks	31	0	2.81	0.87	1.00	5.00	0.35	-0.04
	SE - Apply chem. to everyday tasks	31	0	2.26	0.75	1.00	3.67	-0.06	-0.86
Category 4 - Switched into non STEM	Gender (M=1 and F=2)	10	0	1.50	0.53	1	2	0.00	-2.57
	ACT Composite	9	1	23.44	3.09	17	27	-0.37	-1.59
	ACT Math	9	1	22.78	4.02	16	28	-0.37	-1.05
	ACT Sci-Re	9	1	24.22	3.27	20	31	0.91	1.46
	TP - Math	10	0	81.50	10.55	60	95	-0.94	0.53
	TP - Chemistry	10	0	65.75	5.78	53	73	-1.27	2.55
	TP - Total	10	0	71.01	6.23	62	78	-0.58	-1.22
	OE - Understanding	10	0	1.63	0.36	1.17	2.17	0.27	-1.03
	OE - Performance based tasks	10	0	1.17	0.42	1.00	2.33	2.85	8.32
	OE - Career	10	0	1.62	0.52	1.00	2.50	0.58	-1.14
	OE - Lab tasks	10	0	1.63	0.84	1.00	3.00	0.98	-0.79
	OE - Learner based tasks	10	0	2.67	0.61	1.33	3.33	-0.96	1.67
	SE - Assessment	10	0	1.94	0.46	1.00	2.43	-1.02	0.59
	SE - Interpersonal	10	0	2.30	0.71	1.00	3.33	-0.10	0.23
	SE - Strategies and tasks	10	0	2.47	0.53	2.00	3.67	1.50	2.22
	SE - Low order / recall tasks	10	0	2.40	0.55	1.40	3.00	-0.52	-0.64
	SE - High order tasks	10	0	2.57	0.61	1.67	3.33	-0.26	-1.05
	SE - Apply chem. to everyday tasks	10	0	2.10	0.65	1.00	3.33	0.15	0.87

Selecting the predictors for the persistence model required a different method than obtaining a Pearson-product correlation matrix. Since the outcome was categorical with four levels, determining significant associations between the predictors and the outcome was done using a one-way ANOVA and utilizing the predictors(s) that resulted in significant F-tests. The results of this ANOVA are shown in **Appendix H.** When this method was implemented on the persistence data set, the only significant predictor was SE related to applying problem solving strategies (SE – strategies and tasks). Consequently, this was the only predictor that was used in the development of the persistence model.

Before proceeding with analyses, crosstabs were run to check if the cells in the persistence model were populated. Although multinomial regression is fairly robust against violations of multivariate normality and suited for smaller samples, a check was done regardless. **Tables 8.8** and **8.9** show the results of the crosstab analysis in general and separated by gender respectively.

Table 8.8. Crosstab analysis showing population of each category in the GC I persistence model

		Cross-sectional STEM persistence - GC I start to end				Total
		Stayed in STEM	Switched into STEM	Stayed in non STEM	Switched into non STEM	
Total	Count	385	12	31	10	438
	% of total	87.9%	2.7%	7.1%	2.3%	100.0%

Table 8.9. Crosstab analysis showing population of each category (by gender) in the GC I persistence model

		Cross-sectional STEM persistence - GC I start to end				Total
		Stayed in STEM	Switched into STEM	Stayed in non STEM	Switched into non STEM	
Male	Count	140	3	13	5	161
	% of total	87.0%	1.9%	8.1%	50.0%	100.0%
Female	Count	175	7	12	5	199
	% of total	87.9%	3.5%	6.0%	50.0%	100.0%

Since most students stayed in STEM and the other categories were relatively less populated, there was a possibility of exaggerated effect sizes and estimation of unrealistic coefficients. The analysis was still conducted and the results shown in this chapter have been obtained under this limitation. The addition of more cases in each category in subsequent semesters, tracking students until graduation or narrowing the time point to when students drop the course would offer a more respectable sample size to develop a model and make predictions.

There are two hypotheses of interest in logistic regression:

a) Null: When all the coefficients in the regression equation take on the value of zero.
b) Alternate: The model with predictors currently under consideration is accurate and differs significantly from the null.

The indices that are evaluated to assess fit for the null and full models are shown in **Table 8.10**.

Table 8.10. Model fitting information for GC I cross sectional persistence model

Model	Model Fitting Criteria			Likelihood Ratio Tests		
	AIC	BIC	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	102.144	114.391	96.144			
Final	98.183	122.677	86.183	9.961	3	0.019

The intercept only model (sometimes referred to as the null model) and the final or full model (which includes all the predictors and the intercept) are assessed using two information theory based model fit statistics: The Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). The values for these criteria should be lower for the final model with all predictors, although the BIC tends to be more conservative and results can be mixed (Agresti, 1996). The -2 log likelihood (-2LL) is a likelihood ratio and represents the unexplained variance in the outcome variable. Therefore, smaller the value for this ratio, better the model fit (Agresti, 1996). The GC I full persistence model shows lower AIC and -2LL values than the null model.

The likelihood ratio chi-square test is an alternate test of goodness-of-fit and as with most chi-square tests, prone to inflation as sample size increases. In this case, the model fit is significant χ^2 (3) = 9.961, $p < .05$, which indicates that the full model predicts significantly better than the null model.

Although logistic regression provides pseudo R-square values, these are not displayed here as they are not analogous to R-square values from linear regression and cannot be interpreted in the same way. Information about the utility of the predictor included in the model is obtained using likelihood ratio tests as shown in **Table 8.11**.

Table 8.11. Likelihood ratio tests to indicate importance of predictors in the persistence model

Effect	Model Fitting Criteria			Likelihood Ratio Tests		
	AIC of Reduced Model	BIC of reduced Model	-2 Log Likelihood of reduced Model	Chi-Square	df	Sig.
Intercept	150.008	162.255	144.008	57.825	3	.000
SE-strategies	102.144	114.391	96.144	9.961	3	.019

The statistics in the table above are the same types as those reported for the null and full models in **Table 8.10**. However, in **Table 8.11**, each element of the model is being compared to the full model to make determinations about the inclusivity of the predictor in the full model. In this case, SE -strategies is a significant predictor hypothesized to make a meaningful contribution to the full effect.

The impact this predictor has on the outcome is given by the parameter estimates, shown in **Table 8.12**.

Table 8.12. Parameter estimates for predictors in the GC I persistence model

Persistence category		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp(B)	
								Lower Bound	Upper Bound
1 = Stayed in STEM	Intercept	4.783	1.101	18.859	1	0.000			
	SE - strategies	-0.484	0.429	1.273	1	0.259	0.616	0.266	1.428
2 = Switched into STEM	Intercept	3.143	1.539	4.172	1	0.041			
	SE - strategies	-1.371	0.690	3.945	1	0.047	0.254	0.066	0.982
3 = Stayed in non STEM	Intercept	0.828	1.245	0.442	1	0.506			
	SE - strategies	0.121	0.480	0.064	1	0.800	1.129	0.441	2.890
Reference category is: Switched into non-STEM (4)									

The parameter estimates table shows the logistic coefficient (B) for each predictor variable for each alternative category of the outcome variable. Multinomial logistic regression requires one category to be the reference against which all probabilities and odds are compared. In this case, the reference category is the last one, coded as 4 – switched into non-STEM. The logistic coefficient is the expected amount of change in the logit (what is being predicted) for each unit change in the predictor; it is the odds of membership in the category of the outcome variable which is specified. The closer a logistic coefficient is to zero, the less influence the variable has in predicting the logit (Agresti, 1996). The table also displays the standard error, the Wald statistic, *df*, sig. (*p*-value), Exp(B) and confidence interval for the Exp(B). The Wald test (and associated *p*-value) are used to evaluate whether or not the logistic coefficient is different than zero. The Exp(B) is the odds ratio associated with each predictor. It is expected that predictors which increase the logit will display Exp(B) values greater than 1.0, predictors which do not have an effect on the logit will display an Exp(B) of 1.0 and those which decrease the logit will have Exp(B) less than 1.0 (Agresti, 1996).

In the current model, the only category in which the predictor is significant is 'switching into STEM'. Thus, a decrease in students' average self-efficacy related to using general problem solving strategies makes them .25 times less likely to switch into a STEM major relative to other categories and with all other predictors (if any) staying constant. While this result may not be as impactful given the model's lack of predictive utility, it is meaningful when considering the profile of students who perhaps refrain from entering the physical sciences due to the abstract or mathematical nature of the field, consequently demonstrating low self-efficacy in applying some of the general problem solving strategies required for the tasks in these fields.

Lastly, the ability of this model to correctly classify cases is shown in the classification **Table 8.13**.

Table 8.13. Classification table showing utility of model in categorizing cases

Observed	Predicted				
	Stayed in STEM	Switched into STEM	Stayed in non STEM	Switched into non STEM	Percent Correct
Stayed in STEM	385	0	0	0	100.0%
Switched into STEM	12	0	0	0	0.0%
Stayed in non STEM	31	0	0	0	0.0%
Switched into non STEM	10	0	0	0	0.0%
Overall Percentage	100.0%	0.0%	0.0%	0.0%	87.9%

A perfect model would show values only on the diagonal, indicating correct classification of all cases. The total across the rows represents the number of cases in each category in the actual data while the total down the columns represents the number of cases in each category as classified by the full model (Agresti, 1996). The key piece of information is the overall percentage in the lower right corner which shows the classification accuracy of the current model (with all predictors and the constant) – 87.9%. While this accuracy would ordinarily be considered excellent, it should be viewed cautiously in this case considering the model's lack of differentiation among its predictions and the disproportionate number of students who stayed in a STEM major, thus resulting in exaggerated accuracy values.

Limitations

A typical longitudinal study involves several hundred or several thousand participants who most often represent a national sample. This study was conceptualized to focus on the two-semester sequence of general chemistry courses as a model for persistence in a STEM major since these gateway courses serve as points during which students make choices about staying, switching out of or leaving their intended fields of study. Thus, students who did not take these courses sequentially were excluded from this study. Moreover, as the number of students taking these courses in sequence was minimal and resulted in a small sample size at the end of GC II, the persistence model developed was cross-section in nature and not tested for its predictive utility in an effort to utilize as many students as possible for its development. Thus, interactions between variables and the effects of confounding variables such as socioeconomic status, race and interests were unexplored.

Furthermore, in keeping with the research question addressed in this study, students' majors were classified as either STEM or non-STEM based on the source consulted. As different academic and educational organizations offer varied delineations of what majors constitute a STEM vs. non-STEM category, it is possible some of the majors in this study might have been classified differently.

Despite these limitations, the model developed in this study is the first step towards investigating the affective and cognitive variables that play a role in predicting students' persistence in STEM majors during their enrolment in a single course, during a two-semester gateway course or from a longitudinal perspective.

Conclusion and Implications

The purpose of this study was two-fold: To determine the affective and cognitive factors that impact student performance (on the final exam and in the course) and persistence in a STEM major during their enrolment in GCI. Both research questions in this study were addressed using pre-affective measures – self-efficacy and outcome expectations – and measures of cognitive ability (ACT and TP scores); performance and persistence models were developed using linear and logistic regression respectively.

With regards to the first research question, both measures of cognitive ability (TP total and ACT Composite scores) and self-efficacy related to exam preparation were significant predictors, accounting for 40% of variance in the model based on final exam performance. TP total score, expectations related to success in lab and performance based tasks, and self-efficacy related to applying general strategies accounted for ~ 26% in the model based on performance in the course. As far as the second research question, self-efficacy related to applying strategies was the only influential variable in predicting a student's stability in their intended STEM major while enrolled in GCI. Despite the cross-sectional design of the study, these results show the importance of both performance and affective measures in understanding student accomplishments and stability in an academic major.

These results, while preliminary, bring to light the complexity of issues like STEM persistence and the impact of gender on the persistence differential. Although, from a descriptive standpoint, the percentage of women who stayed in and switched into STEM was higher than those who stayed in non-STEM majors, gender did not play a role in predicting STEM performance or persistence in the regression models. Given the SCCT postulate concerning the mediating role of contextual factors such as support systems on the relationship between gender, career self-efficacy and goals, the direct exclusion of these factors from the models described here could perhaps explain the absence of gender as a significant predictor variable (Lent, Brown & Hackett, 1994). Other studies have shown that the predictive utility of variables in the SCCT framework is not moderated by student's gender (Lent et al., 2005; Lent et al., 2011). While it is possible that gender genuinely does not make a significant contribution to the models in the context of chemistry, the correlational and cross sectional design of these models limits the certainty with which inferences can be made about the temporal ordering of this variable. A longitudinal model with a large sample size and adequately populated persistence categories would allow for better interpretation and generalization across gender subgroups.

Ultimately the objective of any persistence related research, whether in a course or in an intended STEM major, is to increase the number of students who complete a course or a degree with their intended STEM major and to identify students who are 'at risk' for dropping the course, changing to a non-STEM major or leaving prior to attaining a degree and consequently designing interventions to remedy the problems resulting in a lack of persistence. Thus, using a narrower timeline such as students' enrolment in the course after the second exam (the drop date for the course usually ensures the second exam) would address a different persistence outcome; at the same time, broadening the scope of a longitudinal model to include students who chose to enroll in GCII out of sequence would offer a different dataset to examine persistence in STEM majors. This would also allow for the opportunity to interview students to determine what factors play a part in student enrolment in GCII either sequentially or after several semesters. Furthermore, exploring these questions might help understand whether the patterns of STEM-persistence differ for males and females. Given that individual and gender based differences related to STEM domain-knowledge exist even before enrolment in college, female underrepresentation in STEM would be an issue best addressed in high school or perhaps earlier (Ackerman et al., 2013).

CHAPTER 9: OVERALL CONCLUSIONS AND IMPLICATIONS

This chapter summarizes the overall conclusions about understanding students' persistence in STEM majors; the predictive utility of performance models overall and on a much finer scale are discussed. Implications of the findings and limitations for chemical education research are also detailed. Lastly, the potential paths for future research will be discussed based on current findings.

Conclusions

This body of work set out to investigate the impact of performance, self-efficacy and outcome expectations on persistence of students in STEM majors during their enrolment in general chemistry gateway courses. This objective was approached by a) developing a valid and reliable instrument that would provide meaningful measurements for chemistry outcome expectations in first-year chemistry courses, b) adapting a valid and reliable self-efficacy instrument for capturing self-efficacy of students in chemistry courses, c) testing performance models (content based and course performance) to identify predictors that would impact chemistry performance d) integrating performance and affective measures to develop STEM persistence models e) developing a subset instrument to measure these affective constructs on a subtler level and f) utilizing information from these finer measures to identify triggers that would cause a person's affective components to decline thereby placing that person at-risk for leaving or dropping out of a STEM major. Instrument development goals were met through a sequential, exploratory mixed methods design that involved collection of quantitative and qualitative data in combination or in sequence.

Instrument development and psychometric testing

The chemistry outcome expectations survey (COES) resulted in a five-factor solution, which was tested at different time points using CFA to assess its model fit at each point. The necessity to test this model at three time points (GC I pre, GC I post and GC II) was attributed to the ultimate goal of longitudinal data collection. As the longitudinal design would constitute at least three time points during chemistry gateway courses – GC I pre, GC I post and GC II pre and end at GC II post, model fit was tested for each survey's factor structure. The COES factor structure resulted in reasonable to good fit indices at the pre-course time points as opposed to the post-course time point. The CSEAS factor structure also gave reasonable to good fit indices; however, obtaining this structure was an arduous task especially because the surveys used for adaptation had been previously validated. The resulting distinct and meaningful factors, in combination with reasonable fit indices at relevant time points suggested that these instruments are viable measures of each construct for the longitudinal model. Psychometric testing is ongoing for both instruments but preliminary results show that the factors resulting from each survey are meaningful and purport to be measuring unique dimensions of each construct. Additionally, gender differences in several subscales and differences between performance groups on some subscales offer support for the subscales measuring what they are purported to measures. Low performing female students were more positive than male students in their expectations about learner based tasks. While female students displayed lower self-efficacy than male students in subscales relate to interpersonal tasks and applying chemistry everyday tasks, these differences were non-existent at the end of a semester. In addition, low to moderate correlations between self-efficacy, outcome expectations subscales and performance indicators such as the final exam suggest that the surveys were not just alternate measures of academic ability.

Given the pre to post changes occurring in some of these subscales, it was expected that more changes might be occurring during a semester and at key time points such as before or after a testing event. Capturing these affective measures at these points and identifying the triggers would allow for interventions to be developed to offset the lowered affective measure and potentially benefit at-risk students. To that end, a shortened survey was developed by selecting the most meaningful statements from each full length self-efficacy and outcome expectations subscale. This condensed survey was administered at key points throughout a single semester and its subscales were used as predictors to build a performance model to predict scores on each testing event. This baseline or control model accounted for ~34 to 45% of the variance going from exam 1 to exam 3, and resulted in residual averages close to zero at all three time points.

Model testing at key time points

In order to examine if average subscale scores and performance were being lowered prior to or following these testing events, an intervention was developed to assess if affective measures genuinely changed and the predictive performance model residuals were impacted positively or adversely after students were exposed to the intervention. These interventions were packaged as study tools and contained practice problems from past exams along with detailed explanations for solutions to these problems; additionally, an inquiry was conducted into students' problem solving strategies and feedback was offered based on their selected strategies. While these study tools and problem solving strategies in particular were targeted efforts at increasing students' beliefs in their own cognitive and metacognitive strategies, ultimately leading to positive changes in performance and self-efficacy, this did not appear to be the case based on residual analysis and performance group memberships as predicted using discriminant analysis. Data from GC I students who had taken the intervention for exam 2 resulted in negative residuals (~ -2.7%) in comparison to the positive value (~2.1%) that resulted from students in exam 1. This was troubling due to the time point in question being exam 2. As deadlines for dropping a course usually occurred after the second exam, this was an important point to monitor for declines in affective measures and performance. Based on the results from the intervention, it is essential that this data be reexamined and evaluated to identify the potentially egregious predictor(s). It is also quite possible that the second exam was more difficult in the fall term. Techniques such as common item equating would account for these variations and offer a clearer interpretation of the models and associated predictors. Changes in subscale scores were examined, before and after each performance event, using odds ratios. This method was used to assess if the odds of a student moving from a low (before an exam) to a high or better affective group (after the exam) were greater for students who had taken the intervention vs. those who had not. Non-significant odds ratios at all three testing events indicated that the odds of students showing improvements in affective measures was no higher or lower regardless of whether students were in the control or the intervention group. These results indicate that either the interventions might not be operationalized accurately enough to impact students' affective measures or that the interventions are only impacting student performance. However, differences in exams from one semester to the next and students' study habits confounding the way the intervention might be approached (taken as practice with all available resources or as a "quiz" to gauge their preparedness) are factors to consider when examining the utility and efficacy of the study packets.

Methodological limitations

One of the fairly blatant limitations to the subset development and implementation was the inability to model the interaction of each affective measure with time and develop a true growth model by utilizing incomplete datasets as well. Although residuals were assessed at each testing events using a compartmentalized model, the potential of non-independence of residuals was prevalent due to time effects. Although observations were obtained at different times, using two pre measures for exam 1 (start and pre-exam 1) could have resulted in a strong relationship between these variables; while partial regressions and correlations were monitored to make decisions about which time point would offer the most effective predictor between the two listed variables, a time series analysis or modeling techniques would have helped remedy this situation.

Models of persistence mandate accounting for missing data as it is likely not a random occurrence. While missing data were completely excluded from analyses in the models represented in this study, it is essential to address the role of these data not only to better understand the phenomenon of persistence but also to evaluate the impact of nonresponse on the explanatory and predictive power of models described in this body of work.

While discriminant analysis was a unique method to predict group membership, one of the more basic requirements for this technique was the natural occurrence of the groups in the dependent variable as opposed to being created using some criteria such as high vs. low performing groups by using z-scores as a criterion. Given DA's high sensitivity to violations of multivariate techniques, the results of this analysis had to be considered with caution, thus limiting their generality.

The issue of insufficient number of cases was especially problematic in the development of persistence and performance models. While the original intention was to construct a longitudinal model, the filters that were implemented resulted in too few cases for a technique such as multinomial logistic regression. Moreover, as almost all students stayed in a STEM major across a semester, viewing the study using a different lens of persistence such as 'tracking students who enrolled in GC II' would offer a secondary insight into persistence from an enrolment perspective.

Although objective measures such as fixation times, counts and the like were obtained from the usability study and students did appear to be engaging with the problem solving process and ensuing application of strategies for the problems, the strategies did not seem to be functioning as ways to make students more aware of their learning; based on interviews and open response items on the study packet, the strategies seemed to be "in the way" of any "real problem solving" as students had already determined how to alter their preparation and approach to problem solving based on the detailed solutions to each problem. A second phase of interviews where students are solving the problem and going through the entire process of working through a problem and selecting strategies might offer a richer picture of how students integrate the act of detailed problem solving with the use of appropriate strategies.

Implications

The findings reported in this study have several implications for both chemistry teaching and research in chemical education.

Implications for teaching

This study supports some of the hypothesized relationships in social cognitive career theory, by examining relationships between self-efficacy and outcome expectations and their relationships with performance variables. While conceptualizing the affective domain, widely known as emotion, motivation and attitude, is not an easy task, operationalizing these latent variables during teaching is fairly challenging. The findings in this study indicate the complex nature of these variables and the specificity with which they need to be measured. Self-efficacy beliefs in particular can change over the course of a semester, sometimes in highly significant ways; judgments of these beliefs will impact the effort students choose to put into a class, their willingness to persist in an adversarial situation and their willingness to choose to engage in chemistry in the future (Schunk & Pajares, 2001). While these concerns are relatively easy to get trivialized in larger, more impersonal, rigorous college chemistry courses, especially from secondary STEM experiences, and where monitoring students' efficacy beliefs and expectations may not be possible or even practical, efforts can be made to integrate study tools, (relevant to course material) into the course, that will afford students opportunities to reinforce their efforts, receive frequent, focused, task-specific feedback and ultimately create a record of enactive mastery or performance (Margolis & Mccabe, 2006). The findings in this study indicate that task specific and targeted study tools can improve performance on an exam. The results also describe the changes between male and female students on certain dimensions of self-efficacy beliefs and outcome expectations. In particular, low performing female students had more positive expectations than male students about tasks that involved memorization and not understanding concepts; at the same time, female students also displayed lower confidence than male students with regards to interpersonal tasks and applying chemistry to daily tasks. As instructors, it is crucial to understand how affective dimensions can be assessed, how they develop and differ between student subgroups, how to target and offset low affective measures and ultimately impact performance.

Implications for Chemical Education Research

The construct of outcome expectations has been unexplored not just in chemistry but in other domains as well. With existing measures operationalizing this construct in different, and sometimes highly inaccurate ways, the COES offered the first instrument aimed at measuring outcome expectations in chemistry courses. Five subscales measured different dimensions of outcome expectations: Outcome expectations related to understanding chemistry, learner based tasks, performance based tasks, success in lab and career goals. These factors, especially expectations related to career goals, could offer additional information about students' career readiness when examined with seemingly related constructs like career decision making self-efficacy or exploration intentions. Career efficacy and outcome expectation correlated strongly within a group of male college students than with female college students (Betz et al., 1997). These assessments could also be effective in designing career awareness interventions to promote math and science career awareness at educational levels besides college. A measure of outcome expectations would also help fill in the gaps when examining SCCT's hypothesized pathways among its variables. Using this measure in combination with performance and self-efficacy to develop a persistence model would allow for more robust predictions of whether students will persist in their intended STEM majors. While this study used multinomial logistic regression and developed a model by conceptualizing persistence as 'stability in a STEM major', other studies could view persistence from an enrolment standpoint to determine whether students enroll in a sequential course (such as GC II) or persist until the end of the semester in a current course. From a profiling perspective, the findings in this study, utilizing predictive, empirical performance models and odds ratios at key testing events add to the existing work that has been published on identifying at-risk students in chemistry (Chan & Bauer, 2014) and examining the study habits of at-risk students in college general chemistry (Ye et al., 2016).

Future Research

The studies conducted in this project form a small part of an impressive body of work that researchers have been exploring in order to understand persistence in STEM in the context of chemistry and other domains. While two key constructs from SCCT self-efficacy and outcome expectations – constituted the affective measures examined in the models developed in this project, persistence and performance are influenced by other contextual and affective factors that were not accounted or controlled for in these studies. The plethora of subjective measures collected in addition to the survey responses in these studies would provide a much richer profile of students who are enrolling in the course and perhaps allow for early delineation of students based on study habits, types of learners and such. These preliminary profiles would provide a clearer picture of how these groups progress through the semester and what interventions might be appropriate for a group exhibiting lower performance or persistence measures.

The interventions that have been designed contain several general and highly task specific problem solving strategies; evaluating how students engage with these strategies on a deeper level than merely examining them would be highly useful in refining the target time for the intervention as well as the appropriate type of intervention. Subsequent investigations based on gender or major in the course would extensively add to and perhaps clarify the existing and sometimes mixed results on gender based differences in affective research. The efficacy of these interventions can be monitored for the flexibility of students' components of persistence and any resulting changes in predicted career path.

REFERENCES

Ablard, K., & Lipschultz, R. (1998). Self-regulated learning in high-achieving students: Relations to advance reasoning, achievement goals and gender. *Journal of Educational Psychology*, 90(1), 94-101.

Ackerman, P. L., Kanfer, R., & Beier, M.E. (2013). Trait Complex, Cognitive Ability, and Domain Knowledge Predictors of Baccalaureate Success, STEM Persistence, and Gender Differences. *Journal of Educational Psychology,* 105(3), 911-927.

Adelman, C. (2006). *The toolbox revisited: Paths to degree completion from high school through college*. Washington, DC: U.S. Department of Education.

Agresti, A. (1996). An introduction to categorical data analysis. New York, NY: John Wiley & Sons, Inc.

Alper, J. (1993). The pipeline is leaking women all the way along. Science, 260, 409-411.

Astin, A.W. & Astin, H.S. (1993). Undergraduate science education: The impact of different college environments on the educational pipeline in the sciences. Los Angeles, CA: Higher Education Research Institute, U.C.L.A

Astin, A.W., Green, K.C., Korn, W.S. & Schalit, M. (1985). The American Freshman: National norms for fall 1985. Los Angeles, CA: Higher Education Research Institute, U.C.L.A

Azen, R., & Budescu, D.V. (2003). The dominance analysis approach for comparing predictors in multiple regression. *Psychological Methods*, 8, 129-148.

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

Bandura, A. (1989). Perceived self-efficacy in the exercise of personal agency. *The Psychologist: Bulletin of the British Psychological Society*, 2, 411-424.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman

Bandura, A. (2006). *Guide for constructing self-efficacy scales*. In M. Pajares & T.C. Urdan, (Eds.), Self-efficacy beliefs of adolescents (Adolescence and Education, Volume V, ch. 14, pp. 307-337). Information Age Publishing.

Barbera, J., & VandenPlas, J.R. (2011). All assessment materials are not created equal: The myths about instrument development, validity, and reliability. In Bunce, D (ACS Symposium Series – Chapter 11). *Investigating classroom myths through research on teaching and learning*. Washington, DC: American Chemical Society.

Bauer, C.F. (2008). Attitude towards chemistry: A semantic differential instrument for assessing curriculum impacts. *Journal of chemical education*, 85, 1440-1445.

Bean, J.P. (1980). Dropouts and turnover: The synthesis and test of a causal model of student attrition. *Research in Higher Education*, 12, 155-187.

Beavers, A.S., Lounsbury, J.W., Richards, J.K., Huck, S.W., Skolits, G.J., & Esquivel, S.L. (2013). Practical considerations for using exploratory factor analysis in educational research. *Practical Assessment, Research and Evaluation*, 18(6), 1-13. Available online: http://pareonline.net/pdf/v18n6.pdf

Berryman, S.E. (1983). *Who will do Science? Trends, and their causes in minority and female representation among holders of advanced degrees in science and mathematics*, New York, NY: Rockfeller Foundation.

Betz, N.E., & Hackett, G. (1981). The relationship of career-related self-efficacy expectations to perceived career options in college women and men. *Journal of Counseling Psychology*, 28, 399-410.

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, 329-345.

Betz, N.E., Voyten, K.K. (1997). Efficacy and outcome expectations influence career exploration and decidedness. *Career Development Quarterly*, 46(2), 179-189.

Bloom, B., Englehart, M. Furst, E., Hill, W. & Krathwohl, D. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain*. New York, Toronto: Longmans, Green.

Bodner, G.M., & Herron, J.D. (2002). Problem solving in chemistry. In Gilbert, J.K., De Jong, O., Justi, R., Treagust, D.F., & Van Driel, J.H. (Eds.), *Chemical Education: Research-based Practice*. Dordecht: Kluwer Academic Publishers.

Bond, L.A. (1996). Norm- and criterion-referenced testing. *Practical Assessment, Research & Evaluation*, 5(2), 1-3. Available online: http://PAREonline.net/getvn.asp?v=5&n=2.

Bouffard-Bourchard, T., Parent, S., & Larivèe, S. (1991). Influence of self-efficacy on self-regulation and performance among junior and senior high-school aged students. *International Journal of Behavioral Development*, 14, 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*, 12, 271-278.

Brandt, S. A., & Stark, L. W. (1997). Spontaneous Eye Movements During Visual Imagery Reflect the Content of the Visual Scene. *Journal of Cognitive Neuroscience 9*(1), pp. 27-38

Brooks, L. & Betz, N.E. (1990). Utility of expectancy theory in predicting occupational choices in college students. *Journal of Counseling Psychology*, 37, 57-64.

Brown, S.D., Tramayne, S., Hoxha, D., Telander, K., Fan, X., & Lent, R.W. (2008). Social cognitive predictors of college students' academic performance and persistence: A meta-analytic path analysis. *Journal of Vocational Behavior*, 72, 298-308.

Browne, M.W. (2000). Cross-validation methods. *Journal of Mathematical Psychology*, 44, 108-132.

Business-Higher Education Forum. (2010). Increasing the number of STEM graduates: Insights from the U.S. STEM Education & Modeling Project. Washington, D.C.: Author.

Cabrera, A.F., Castañeda, M.B., Nora, A. & Hengstler, D. (1992). The convergence between two theories of college persistence. *Research in Higher Education*, *63*(2), 144-164.

Cates, G.L., & Rhymer, K.N. (2003). Examining the relationship between mathematics anxiety and mathematics performance: An instructional hierarchy perspective. *Journal of Behavioral Education*, 12, 23-34.

Chan, J.Y.K., & Bauer, C.F. (2014). Identifying at-risk students in general chemistry via cluster analysis of affective characteristics. *Journal of Chemical Education*, 91(9), 1417-1425.

Chemers, M.M., Hu, L.-T., & Garcia, B.F. (2001). Academic self-efficacy and first-year college student performance and adjustment. *Journal of Educational Psychology*, 93(1), 55-64.

Child, D. (2006). *The essentials of factor analysis* (3rd ed.). New York, NY: Continuum International Publishing Group.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd edition). Hillsdale, NJ: Lawrence Erlbaum Associates.

Coll, R.K., Dalgety, J. & Salter, D. (2002). The Development of the Chemistry Attitudes and Experiences Questionnaire (CAEQ). *Chemistry Education Research and Practice in Europe*, 3(1), 19-32.

Comrey, A.L. (1985). A method for removing outliers to improve factor analytic results. *Multivariate Behavioral Research*, 20, 273-281.

Cook, R.D. & Weisberg, S. (1999). *Applied Regression Including Computing and Graphics*.

Hoboken NJ: Wiley.

Copas, J.B. (1987). Cross-validation shrinkage of regression parameters. *Journal of the Royal Statistical Society. Series B (Methodological)*, 49(2), 175-183.

Costello, A.B., & Osborne, J.W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment Research & Evaluation*, 10(7), 1-9. Available online: http://www.pareonline.net/pdf/v10n7.pdf

Creswell, J.W. (2003). *Research design: Qualitative, quantitative, and mixed methods approaches* (2nd ed.). Thousand Oaks, CA: Sage

Danaher, K. & Crandall, C. S. (2008). Stereotype threat in applied settings re-examined. *Journal of Applied Social Psychology,* 38(6), 1639-1655.

Dave, R.H. (1970). *Psychomotor levels in Developing and Writing Behavioral Objectives*, pp. 20-21. R.J. Armstrong, ed. Tucson, AZ: Educational Innovators Press.

Dawis, R.V. (1987). Scale construction. *Journal of Counseling Psychology*, 34, 481-489.

Dey, E.L., Astin, A.W. & Korn, W.S. (1991). The American Freshman: Twenty-five year trends. Los Angeles, CA: *Higher Education Research Institute*, U.C.L.A

Drost, E.A. (2011). Validity and reliability in social science research. *Education Research and Perspectives*, 38(1), 105-123.

Dumas, J. S. and Redish, J. C. (1999). *A Practical Guide to Usability Testing*. Portland, OR: Intellect Ltd.

Eddy, R.M. (2000). Chemophobia in the College Classroom: Extent, Sources and Student Characteristics. *Journal of Chemical Education*, 77(4), 514-517.

Elliot, R., Strenta, A.C., Adair, R., Matier, M., & Scott, J. (1996). The role of ethnicity in choosing and leaving science in highly selective institutions. *Research in Higher Education*, 37(6), 681-709.

Everitt, B.S., Landau, S., Leese, M., & Stahl, D. (2001). Cluster analysis (4th ed.). New York, NY: Oxford University Press Inc.

Felder, R.M., Felder, G.N., Mauney, M., Hamrin, C.E., & Dietz, E.J. (1995). A longitudinal study of engineering student performance and retention: Gender differences in student performance and attitudes. *Journal of Engineering Education*, 84(2), 151-163.

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

Field, A. (2009). *Discovering Statistics Using SPSS: Introducing Statistical Methods* (3rd ed.). Thousand Oaks, CA: Sage Publications.

Fouad, N.A., & Smith, P.L. (1996). A test of a social cognitive model for middle school students: Math and Science. *Journal of Counseling Psychology*, 43(3), 338-346.

Fouad, N.A., & Guillen, A. (2006). Outcome expectations: Looking to the past and potential future. *Journal of career assessment*, 14, 130-142.

Fouad, N.A., Hackett, G., Smith, P.L., Kantamneni, N., Fitzpatrick, M., Haag, S., & Spencer, D. (2010). Barriers and supports for continuing in mathematics and science: Gender and educational level differences. *Journal of Vocational Behavior*, 77, 361-373.

George-Jackson, C.E. (2011). STEM switching: Examining departures of undergraduate women in STEM fields. *Journal of women and minorities in science and engineering*, 17(2), 149-171.

Germann, P.J. (1988). Development of the attitude toward science in school assessment and its use to investigate the relationship between science achievement and attitude toward science in school. *Journal of Research in Science Teaching*, 25, 689-703.

Gorman, B.S., & Primavera, L.H. (1983). The complementary uses of cluster and factor analysis methods. *The Journal of Experimental Education*, 51(4), 165-168.

Grove, N., & Bretz, S.L. (2007). CHEMX: An instrument to assess students' cognitive expectations for learning chemistry. *Journal of Chemical Education*, 84(9), 1524-1529.

Grunert, M.L. & Bodner, G.M. (2011). Finding fulfillment: women's self-efficacy beliefs and career choices in chemistry. *Chemistry education research and practice*, 12, 420-426.

Guo, Y., Hastie, T., & Tibshirani, R. (2007). Regularized discriminant analysis and its application in microarray. *Biostatistics*, 8(1), 86-100.

Hackett, G. (1985). The 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, 47-56.

Hackett, G., & Betz, N.E. (1989). An exploration of mathematics self-efficacy, mathematics performance. Correspondence. *Journal for research in mathematics education*, 20, 261-273.

Hackett, G., Betz, N.E., Casas, J.M., & Rocha-Singh, I. (1992). Gender, ethnicity and social cognitive factors predicting the academic achievement of students in engineering. *Journal of Counseling Psychology*, 39, 527-538.

Hackett, G. (1995). Self-efficacy in career choice and development. In Bandura (Ed.), *Self-*

efficacy in changing societies (pp. 232-258). New York: Cambridge University Press.

Hall, R.M. & Sandler, B.R. (1982). *The Classroom Climate: A Chilly One for Women.* Washington DC: Project on the Status and Education of Women, Association of American Colleges.

Hall, J.M., & Ponton, M.K. (2005). Mathematics self-efficacy of college freshman. *Journal of Developmental Education*, 28(3), 26-32.

Hampton, N.Z. & Mason, E. (2003). Learning disabilities, gender, sources of self-efficacy, self-efficacy beliefs, and academic achievement in high school students. *Journal of School Psychology*, 41, 101-112.

Harrow, A. (1972). *A Taxonomy of Psychomotor Domain: A Guide for Developing Behavioral Objectives.* New York, NY: David McKay Co., Inc.

Henson, R., & Roberts, J. (2006). Use of exploratory factor analysis in published research: Common errors and some comment on improved practice. *Educational and Psychological Measurement*, 66, 393-416.

Horn, J.L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30, 179-185.

Horn, L., & Kojaku, L.K. (2001). *High school academic curriculum and the persistence path through college* (NCES 2001-163). Washington, DC: U.S. Department of Education, National Center for Education Statistics.

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

Hughes, R. (2011). Are the predictors of women's persistence in STEM painting the full picture? A series of comparative case studies. *International Journal of Gender, Science and Technology*, 3(3), 547-570.

Inda, M., Rodríguez, C., & Peña, J. V. (2013). Gender differences in applying social cognitive career theory in engineering students. *Journal of Vocational Behavior*, 83, 346-355.

Johnson, J.W., & LeBreton, J.M. (2004). History and use of relative importance indices in organizational research. *Organizational Research Methods*, 7, 238-257.

Josephson, S., & Holmes, M.E. (2002). Visual attention to repeated Internet images: Testing the scan path theory on the world wide web. In *Proceedings of the Eye Tracking Research & Applications Symposium 2002* (pp. 43-49). New York, ACM.

Just, M.A. and Carpenter, P.A. (1980) A Theory of Reading: From Eye Fixations to Comprehension. *Psychological Review*, 87, 329-354.

Klecka, W.R. (1980). *Discriminant Analysis* (Sage University Paper series on Quantitative Applications in the Social Sciences, No. 07-019). Thousand Oaks, CA: Sage

Kline, P. (2002). *An easy guide to factor analysis*. London: Routledge.

Kort, F. (1973). Regression analysis and Discriminant analysis: An application of R.A.Fisher's theorem to data in political science. The American Political Science Review, 67(2), 555-559.

Krathwohl, D.R., Bloom, B.S. & Masia, B.B. (1973). *Taxonomy of Educational Objectives, the Classification of Educational Goals. Handbook II: Affective Domain*, New York, NY: David McKay Co., Inc.

Laeng, B., Sirois, S., & Gredeback, G. (2012). Pupillometry: A window to the preconscious? *Perspectives on Psychological Science*, 7, 18 –27.

Lent, R.W., Brown, S.D., & Larkin, K.C. (1984). Relation of self-efficacy to academic achievement and persistence. *Journal of Counseling Psychology*, 31, 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, 265-269.

Lent, R.W., Lopez, F.G., & Bieschke, K. (1991). Mathematics self-efficacy: Sources and relation to science-based career choice. *Journal of counseling psychology*, 34, 424-430.

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, 79-122.

Lent, R.W., Brown, S.D., Sheu, H., Schmidt, J., Brenner, B.R., Gloster, C.S., Wilkins, G., Schmidt, L., Lyons, H., & Treistman, D. (2005). Social cognitive predictors of academic interests and goals in engineering: Utility for women and students at historically Black universities. *Journal of Counseling Psychology*, 52, 84-92.

Lent, R.W., & Brown, S.D. (2006). On conceptualizing and assessing social cognitive constructs in career research: A measurement guide. *Journal of Career Assessment*, 14, 12-35.

Lent, R.W., Lopez Jr, A.M., Lopez, F.G., & Sheu, H. (2008). Social cognitive career theory and the prediction of interests and choice goals in the computing disciplines. *Journal of Vocational Behavior*, 73, 53-62.

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, 328-335.

Lynch, D.J., & Trujillo, H. (2010). Motivational beliefs and learning strategies in organic chemistry. *International Journal of Science and Mathematics Education*, 9, 1351-1365.

Lent, R.W., Lopez, A.M., Sheu, H., & Lopez, A. (2011). Social cognitive predictors of the interest and choices of computing majors: Applicability to underrepresented students. *Journal of Vocational Behavior*, 78, 184-192.

Maltese, A.V., & Tai, R.H. (2010). Eyeballs in the fridge: Sources of early interest in science. *International Journal of Science Education*, 32(5), 669-685.

Maltese, A.V. & Tai, R.H. (2011). Pipeline persistence: Examining the association of educational experiences with earned degrees in STEM among U.S. students. *Science Education*, 95, 877-907.

Margolis, H., & Mccabe, P.P. (2006). Improving self-efficacy and motivation: What to do, what to say. *Intervention in School and Clinic*, 41(4), 218-227.

Marra, R.M., Schuurman, M., Moore, C., & Bogue, B. (2005). Women engineering students' self-efficacy beliefs – The longitudinal picture. *Proceedings of the annual meeting of the American Society for Engineering Education Annual Conference*, 12 – 15 June, Portland, OR.

Marra, R.M., & Bogue, B. (2006). Women engineering students' self-efficacy – A longitudinal Multi-institution study. *Proceedings of the Women in Engineering Programs and Advocates Network (WEPAN).*

Maruyama, G.M. (1998). *Basis of structural equation modeling*. Thousand Oaks, CA: Sage.

Mau, W.-C. (2003). Factors that influence persistence in science and engineering career aspirations. *Career Development Quarterly*, 51(3), 234-243.

May, G.S., & Chubin, D.E. (2003). A retrospective on undergraduate engineering success for underrepresented minority students. *Journal of engineering education*, 92(1), 27-40.

McHugh, M.L. (2009). The odds ratio: Calculation, usage, and interpretation. *Biochemia Medica*, 19(2), 120-126.

McWhirter, E.H. (1997). Perceived barriers to education and career: Ethnic and gender differences. *Journal of Vocational Behavior*, 50, 124-140.

McWhirter, E., Crothers, M., & Rasheed, S. (2000). The effects of high school career education on social cognitive variables. *Journal of Counseling Psychology*, 47, 330-341.

Menard, S. (1995). *Applied logistic regression analysis*. Thousand Oaks, CA:Sage

Miller, D.I. & Wai, J. (2015). The bachelor's to Ph.D. STEM pipeline no longer leaks more women to men: a 30-year analysis. *Frontiers in Psychology*, 6(37), 1-10.

Mishel, M.H. (1998). Methodological studies: Instrument development. In P.J. Brink & M.J. Wood (Eds.), *Advanced design in nursing research* (2nd ed., pp. 235-282). Thousand Oaks, CA: Sage.

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, 30-38.

Napier, J.D., & Riley, J.P. (1985). Relationship between affective determinants and achievement in science for seventeen-year-olds. *Journal of Research in Science Teaching*, 22, 365-383.

National Education Goals Panel. (1993). *The National Education Goals Report, Volume One: The National Report 1993*. Washington, DC: U.S. Government Printing Office.

National Science Foundation. (2010). Integrated postsecondary education data system completions survey [data file]. From WebCASPAR: https://caspar.nsf.gov/.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

O'Connor, B.P. (2000). SPSS and SAS programs for determining the number of components using Parallel Analysis and Velicer's MAP Test. *Behavior Research Methods, Instruments and Computers*, 32(3), 396-402.

Osborne, J.W. (2000). Prediction in Multiple Regression. *Practical Assessment Research & Education*, 7(2), 1-6. Available online: http://www.pareonline.net/pdf/v7n2.pdf

Osborne, J.W. (2015). What is rotating in exploratory factor analysis? *Practical Assessment, Research & Evaluation*, 20(2). Available online: http://www.pareonline.net/pdf/v20n2.pdf

Pace, C.R., & Kuh, G.D. (1998). *College Student Experiences Questionnaire* (4th ed.). Bloomington, IN: Indiana University.

Pajares, F., & Miller, M.D. (1994). The role of self-efficacy and self-concept beliefs in mathematical problem-solving: A path analysis. *Journal of Educational Psychology*, 86, 193-203.

Pajares, F. (1996). Self-efficacy beliefs in academic settings. Review of Educational Research, 66(4), 543-578.

Pajares, F. (1997). Current directions in self-efficacy research. In M.Maehr & P.R.Pintrich (Eds.), Advances in motivation and achievement (Volume 10, pp. 1-49). Greenwich, CT: JAI Press.

Pajares, F., & Schunk, D.H. (2001). Self-beliefs and school success: Self-efficacy, self-concept and school achievement. In R.Riding & S.Rayner (Eds.), *Perception* (pp. 239-266). London: Ablex Publishing.

Palmer, S. E. (2002). *Vision science: Photons to phenomenology*. Cambridge, MA: MIT Press.

Patil, V.H., Singh, S.N., Mishra, S., & Donovan, T. (2007). Parallel analysis engine to aid determining number of factors to retain [Computer software]. Available from http://smishra.faculty.ku.edu/parallelengine.htm

Patil, V.H., Singh, S.N., Mishra, S., & Donovan, T. (2008). Efficient theory development and factor retention criterial: A case for abandoning the 'eigenvalue greater than one' criterion. *Journal of Business Research*, 61(2), 162-170.

Patrick, H., & Hicks, L. (1997). Relations of perceived self-efficacy and social goal pursuit. *Journal of Early Adolescence*, 17, 109-129.

Pavelich, M.J. (1982). Using general chemistry to promote the higher level thinking abilities. *Journal of Chemical Education*, 59(9), 721-724.

PCAST. (2012). President's Council of Advisors on Science and Technology. Prepare and inspire: K-12 education in science, technology, engineering and math (STEM) for America's future. From PCAST: http://www.whitehouse.gov/ostp/pcast.

Pedhazur, E.J. (1997). *Multiple regression in behavioral research*. Harcourt Brace: Orlando, FL.

Pett, M.A., Lackey, N.R. & Sullivan, J.J. (2003). *Making Sense of Factor Analysis: The Use of Factor Analysis for Instrument Development in Health Care Research*. Thousand Oaks, CA: Sage.

Pintrich, P.R., & De Groot, E.V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, 82, 33-40.

Pintrich, P.R., & Schunk, D.H. (1995). *Motivation in education: Theory, research and applications*. Englewood Cliffs, NJ: Prentice Hall.

Plake, B.S., & Parker, C.S. (1982). The development and validation of a revised version of the Mathematics Anxiety Rating Scale. *Educational and Psychological Measurement*, 42(2), 551-557.

Polya, G. (1957). *How to solve it*. Garden City, NY: Doubleday and Co. Inc.

Poole, A., & Ball, L. J. (2006). Eye tracking in human-computer interaction and usability research: Current status and future prospects. In C. Ghaoui (Ed.), *Encyclopedia of human computer interaction*, 211–219. Pennsylvania, PA: Idea Group.

Pritchard, M.E, & Wilson, G.S. (2003). Using emotional and social factors to predict student success. Journal of College Student Development, 44(1), 18-28.

Rattray, J., & Jones, M.C. (2007). Essential elements of questionnaire design and development. *Journal of Clinical Nursing*, 16, 234-243.

Reason, R.D. (2009). An exam of persistence research through the lens of a comprehensive conceptual framework. *Journal of College Student Development*, *50*(6), 659-682.

Redish, E.F., Saul., J.M., & Steinberg, R.N. (1998). Student expectations in introductory physics. *American Journal of Physics*, 66(3), 212-224.

Richardson, R., & Suinn, R.N. (1972). The mathematics anxiety rating scale: Psychometric data. *Journal of Counseling Psychology*, 19, 551-554.

Riggs, M.L., Warka, J., Babasa, B., Betancourt, R., & Hooker, S. (1994). Development and validation of self-efficacy and outcome expectancy scales for job related applications. *Educational and Psychological Measurement*, 54, 793-802.

Robbins, S.B., Lauver, K., Le, H., Davis, D., & Langley, R. (2004). Do psychosocial and study skill factors predict college outcomes? A meta-analysis. *Psychological Bulletin*, 130, 261-288.

Ruspini, E. (2002). *Introduction to Longitudinal Research*. New York, NY: Routledge.

Schmitt, T.A. (2011). Current methodological considerations in exploratory and confirmatory factor analysis. *Journal of Psychoeducational Assessment*, 29(4), 304-321.

Schunk, D.H. (1981). Modeling and attributional feedback effects on children's achievement: A self-efficacy analysis. *Journal of Educational Psychology*, 74, 93-105.

Schunk, D. H., & Hanson, A. R. (1985). Peer models: Influence on children's self-efficacy and achievement. *Journal of Educational Psychology*, 77,313-322.

Schunk, D.H. (1987). Peer models and children's behavioral change. *Review of Educational Research*, 57, 149-174.

Schunk, D. H. (1989). Self-efficacy and achievement behaviors. *Educational Psychology Review*, 1, 173-208.

Schunk, D.H. (1991). Self-efficacy and Academic Motivation. *Educational Psychologist*, 26, 207-231.

Schunk, D. H. (1995). Self-efficacy, motivation, and performance. *Journal of Applied Sport Psychology*, 7(2), 112-137.

Schunk, D.H. (1996). Self-efficacy for learning and performance. Paper presented at the meeting of the American Educational Research Association, New York.

Seymour, E. & Hewitt, N.M. (1997). *Talking about leaving: Why undergraduates leave the sciences*, Boulder, CO: Westview.

Shell, D.F., Colvin, C., & Bruning, R.H. (1995). Self-efficacy, attributions, and outcome expectancy mechanisms in reading and writing achievement: Grade level and achievement-level differences. *Journal of Educational Psychology*, 87, 386-398

Simpson, E.J. (1972). *The Classification of Educational Objectives in the Psychomotor Domain.* Washington, DC: Gryphon House.

Smist, J.M. (1992, October). Science self-efficacy among high school students. *Paper presented at the annual meeting of the Northeastern Educational Research Association, Ellenville, NY.*

Smith, P.L., & Fouad, N.A. (1999). Subject-matter specificity of self-efficacy, outcome expectations, interests and goals: Implications for the social-cognitive model. *Journal of Counseling Psychology*, 46, 461-471.

Soofi, E.S., Retzer, J.J., & Yasai-Ardekani, M. (2000). A framework for measuring the importance of variables with applications to management research and decision models. *Decision Sciences*, 31, 595-625.

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology,* 35(1), 4-28.

Springer, S.H., Larson, L.M., Tilley, B.P., Gasser, C.E., & Quinn, A.C. (2001). The development of an Educational and Career Outcome Expectancy Scale. *Paper presented at the 109th Annual Meeting of the American Psychological Association* (San Francisco, CA).

Steele, C.M. & Aronson, J. (1995). Stereotype threat and the Intellectual Test Performance of African Americans. *Journal of Personality and Social Psychology*, 69(5), 797-811.

Steele, C.M. (1997). A Threat in the Air: How Stereotypes Shape Intellectual Identity and Performance. *American Psychologist*, 52(6), 613-629.

STEM Designated Degree program list., May 2016 https://www.ice.gov/sites/default/files/documents/Document/2016/stem-list.pdf

Stevens, J. (2002*). Applied multivariate statistics for the social sciences* (4th Ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Tabachnick, B.G., & Fidell, L.S. (2001). *Using multivariate statistics* (4th ed.). Boston: Allyn & Bacon.

Tanaka, J.S. (1993). Multifaceted conceptions of fit in structural equation models. In K.A. Bollen, & J.S. Long (eds.). *Testing structural equation models*. Newbury Park, CA: Sage.

Tannenbaum, C., Greaves, L., & Graham, I.D. (2016). Why sex and gender matter in implementation research. *BMC Medical Research Methodology*, 16(1), 1-9.

Terenzini, P.T. & Reason, R.D. (2005). Parsing the first year of college: Rethinking the effects of college on students. *Paper presented at the Annual Conference of the Association for the Study of Higher Education, Philadelphia, PA.*

Thompson, B., & Daniel, L.G. (1996). Factor analytic evidence for the construct validity of scores: A historical overview and some guidelines. *Educational and Psychological Measurement*, 56, 197-208.

Thorndike, R.M. (1978). *Correlational procedures for research*. New York, NY: Gardner. Miles, J., & Banyard, P. (2007). *Understanding and using statistics in psychology: A practical introduction*. London: Sage

Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. *Review of Educational Research*, 45, 89-125.

Toven-Lindsey, B., Levis-Fitzgerald, M., Barber, P.H., & Hasson, T. (2015). Increasing Persistence in Undergraduate Science Majors: A Model for Institutional Support of Underrepresented Students. *CBE-Life Sciences Education*, 14, 1-12.

Towns, M.H. (2008). Using inferential statistics to answer quantitative chemical education research questions. In D.M. Bunce & R.S. Cole (Eds.), *Nuts and bolts of chemical education research* (pp. 135-148). Washington, DC: American Chemical Society.

Trochim, W.M.K. (2000). The Research Methods Knowledge Base (2nd ed.). Cincinnati, OH: Atomic Dog Publishing.

Trusty, J. (2002). Effects of high school course-taking and other variables on choice of science and mathematics college majors. *Journal of Counseling and Development*, 80, 464-474.

Tyson, W., Lee, R., Borman, K.M., & Hanson, M.A. (2007). Science, technology, engineering and mathematics (STEM) pathways: High school science and math coursework and postsecondary degree attainment. *Journal of Education for Students Placed at Risk*, 12(3), 243-270.

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.

Van den Burg, W., & Lewis, C. (1988). Some properties of two measures of multivariate association. *Psychometrika*, 53, 109-122.

Waldrop, M.M. (2015). Why we are teaching science wrong, and how to make it right. *Nature*, 523, 272-274.

Waller, B. (2005). Math interest and choice intentions of non-traditional African-American college students. *Journal of Vocational Behavior*, 68, 538-547.

Waltz, C.F., Strickland, O.L. & Lenz, E.R. (1991). *Measurement in nursing research. the Study of Higher Education, Philadelphia, PA.*
Philadelphia: F.A. Davis.

Ware, N.C., & Lee, V.E. (1988). Sex differences in choice of college science majors. *American Educational Research Journal*, 25(4), 593–614.

Williams, W. M. & Ceci, S. J. (2007) Introduction: Striving for Perspective in the Debate on Women in Science, In S. J. Ceci and W. M. Williams, Eds., *Why Aren't More Women in Science? Top Researchers Debate the Evidence*, Washington DC: American Psychological Association.

Xu, X., & Lewis, J.E. (2011). Refinement of a chemistry attitude measure for college students. Journal of Chemical Education, 88, 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, 188-200.

Ye, L., Shuniak, C., Oueini, R., Robert, J., & Lewis, S. (2016). Can they succeed? Exploring at-risk students' study habits in college general chemistry. *Chemistry Education Research and Practice*, 17, 878-892.

Yong, A.G., & Pearce, S. (2013). A beginner's guide to factor analysis: Focusing on exploratory factor analysis. *Tutorials in Quantitative Methods for Psychology*, 9(2), 79-94.

Zaracova, A., Lynch, S.M., & Espenshade, T.J. (2005). Self-efficacy, stress, and academic success in college. *Research in Higher Education*, 46(6), 677-706.

Zimmerman, B.J., & Martinez-Pons, M. (1990). Student differences in self-regulated learning: Relating grade, sex, and giftedness to self-efficacy and strategy use. *Journal of Educational Psychology*, 82, 51-59.

Zimmerman, B., Bandura, A., & Martinez-Pons, M. (1992). Self-motivation for academic achievement: The role of self-efficacy beliefs and personal goal setting. *American Educational Research Journal*, 29, 663-676.

Zimmerman, B.J. (1999). Commentary: Toward a cyclically interactive view of self-regulated learning. *International Journal of Educational Research*, 31, 545-551.

Zimmerman, B.J. (2000). Self-efficacy: An essential motive to learn. *Contemporary*

Educational Psychology, 25, 82-91.

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, 1081-1094.

APPENDICES

APPENDIX A:

SAS code used to conduct confirmatory factor analysis (CFA)

```
PROC IMPORT datafile='C:\Users\Shalini\Desktop\filename' dbms=tab OUT=Shalini.Subscale
replace;
GETNAMES=YES;
DATAROW=2;
RUN;
PROC PRINT;
RUN;

proc calis data=Shalini.Subscale CORR RESIDUAL modification;

factor
F1 ---> OE18,
F1 ---> OE16,
F1 ---> OE14,
F1 ---> OE9,
F1 ---> OE19,
F1 ---> OE20,

F2 ---> OE22,
F2 ---> OE1,
F2 ---> OE24,
F2 ---> OE12,


F3 ---> OE15,
F3 ---> OE3,
F3 ---> OE6,
F3 ---> OE7,


F4 ---> OE4,
F4 ---> OE25,
F4 ---> OE11,


F5 ---> OE10,
F5 ---> OE23,
F5 ---> OE8;


pvar
F1 F2 F3 F4 F5 = 5*1.;
run;
```

APPENDIX B:

IRB Consent form – class-wide data collection

CONSENT TO PARTICIPATE IN A RESEARCH STUDY

IMPACT on STUDENT LEARNING in UNDERGRADUATE SCIENCE COURSES

Study to be conducted at:	University of Wisconsin - Milwaukee, Chemistry and Biochemistry Department	
Principal Investigators:	Kristen Murphy	229-4468
IRB Approval date:	06/16/2015	IRB #: 14.404

INFORMATION:
You are invited to participate in a research study. The Institutional Review Board (IRB) of the University of Wisconsin – Milwaukee (UWM) has reviewed this study for the protection of the rights of human subjects in research studies, in accordance with federal and state regulations. However, before you choose to be a research participant, it is important that you read the following information and ask as many questions as necessary to be sure that you understand what your participation will involve. Your signature on this consent form will acknowledge that you received all of the following information and explanations from the principal investigator (or his/her designated representative), and have been given an opportunity to discuss your questions and concerns with the principal investigator or a co-investigator. Additionally, should you have any questions regarding your rights as a human participant, please do not hesitate to contact Institutional Review Board at 414-229-3173.

PURPOSE:
This study involves research into effective strategies for improving learning in chemistry. Approximately 1500 students per semester will be involved in this research.

PROCEDURES:
If you agree to participate in this study, you are giving permission to the researchers to use your class data (exam and quiz responses, laboratory reports, informal assessment questions, survey responses, practice exam responses, worksheet responses, and final grade in the course) and your demographic data (gender, year in school, ACT scores and subscores, placement test scores, major or intended major, minor, high school, year of high school graduation, high school GPA and class rank) for research purposes. Although your name will be linked to your data, all reports and results will report aggregated data only. Your personal identifiers will be removed from the dataset at the end of the study and your deidentified data will be kept indefinitely. Your classroom and demographic data will only be used for research purposes if you agree to participate in this study.

POSSIBLE RISKS:
This project is minimal risk research. Any statements or actions on your part will not be identified by your name or any other identifier to anyone outside the project. This is accomplished through the use of secure computer facilities (password protected) for digital data or kept in a locked cabinet. Only the research team has access to this material. Your participation in this project will be held in confidence, however results of the project may be published. Any results from this project will not contain information by which you may be identified.

EXCLUSION REQUIREMENTS:
Students under 18 years of age will not participate in this research study.

POTENTIAL BENEFITS:
The potential benefits from this research include: improved problem solving skills and improved content mastery. **You may be given extra credit points for participating in specific components of this study.** *It is not possible to predict whether or not any personal benefit will result from your participation in this study. You understand that the information that is obtained from this study may be used scientifically and may be helpful to others.*

Participant's Initials ________

Revised 08/26/2015

Page 1 of 2

APPENDIX C:

IRB Consent form – think aloud / interviews

CONSENT TO PARTICIPATE IN A RESEARCH STUDY

IMPACT on STUDENT LEARNING in UNDERGRADUATE SCIENCE COURSES

Study to be conducted at:	University of Wisconsin - Milwaukee, Chemistry and Biochemistry Department
Principal Investigators:	Kristen Murphy 229-4468
IRB Approval date:	06/16/2015 IRB #: 14.404

INFORMATION:
You are invited to participate in a research study. The Institutional Review Board (IRB) of the University of Wisconsin – Milwaukee (UWM) has reviewed this study for the protection of the rights of human subjects in research studies, in accordance with federal and state regulations. However, before you choose to be a research participant, it is important that you read the following information and ask as many questions as necessary to be sure that you understand what your participation will involve. Your signature on this consent form will acknowledge that you received all of the following information and explanations from the principal investigator (or his/her designated representative), and have been given an opportunity to discuss your questions and concerns with the principal investigator or a co-investigator. Additionally, should you have any questions regarding your rights as a human participant, please do not hesitate to contact Institutional Review Board at 414-229-3173.

PURPOSE:
This study involves research into effective strategies for improving learning in chemistry. Approximately a maximum 80 students per semester will be involved in this research.

PROCEDURES:
You will have an opportunity to volunteer to participate in a thirty to sixty-minute interview that may be videotaped. Additionally, you may have the opportunity to participate in an activity that would involve working problems at a computer that will track you eye movement in order to measure problem-solving strategies. Additionally, we will collect demographic data about you. Participation in the interviews is voluntary and not a required component of the course.

POSSIBLE RISKS:
This project is minimal risk research. Any statements or actions on your part will not be identified by your name or any other identifier to anyone outside the project, and your participation in this project will be held in confidence, however results of the project may be published. Any results from this project will not contain information by which you may be identified.

EXCLUSION REQUIREMENTS:
Students under 18 years of age will not participate in this research study.

POTENTIAL BENEFITS:
The potential benefits from this research include: improved problem solving skills and improved content mastery. You may be given financial compensation in the form of a gift card to the UWM Bookstore or a study guide for participating in specific components of this study. *It is not possible to predict whether or not any personal benefit will result from your participation in this study. You understand that the information that is obtained from this study may be used scientifically and may be helpful to others.*

VOLUNTARY PARTICIPATION:
Participation in this study is voluntary. You may refuse to participate or withdraw from this study at any

Participant's Initials ________

time. We may collect demographic data about you. This can include gender, year in school, ACT scores and subscores, placement test scores, major or intended major, minor, high school, year of high school graduation, high school GPA and class rank. Research data, however, will not be collected on students who elect not to participate. *If you elect not to participate or withdraw from the study at any time, you will not be penalized or lose any benefits associated with the viewing the images generated by the instrumentation and your decision will not affect your relationship with this institution or your standing in the chemistry course. If you elect not to participate in this study, you will still be eligible for possible extra credit points.*

The investigator may withdraw you from the study at any time. If this is done it will not affect your grade in your chemistry course.

You will be informed of any significant new information regarding this study that may affect your willingness to continue in this study.

You would be investing approximately one (1) hour of your time in each opportunity for participation in this study.

CONFIDENTIALITY:
The records of your participation are confidential. The investigator will maintain your information, and this information may be kept on a computer. Study information or data may be examined by the Institutional Review Board of the University of Wisconsin - Milwaukee and various federal regulatory agencies. This study may result in scientific presentations and publications, but steps will be taken to ensure you are not identified by name.

QUESTIONS:
For more information concerning this study and research-related risks or injuries, you may contact the Principal Investigator (see first page for identifying information). You may also contact a representative of the Institutional Review Board of UWM for information regarding rights of participants involved in a research study.

CONSENT:
I have been given an opportunity to ask questions about this study; answers to such questions (if any) have been satisfactory.

In consideration of all of the above, I give my consent to participate in this research study. I acknowledge receipt of a copy of this informed consent statement.

PARTICIPANT'S SIGNATURE: ____ DATE________________

PARTICIPANT'S PRINTED NAME: ____

Please sign here if you choose NOT to participate: ____

Principal Investigator: Kristen Murphy 414-229-4468

Participant's Initials ________

APPENDIX D:

Fall 2012 Self-efficacy standalone survey

How confident are you about:	Not Confident at all	Barely Confident	Neutral	Fairly Confident	Totally Confident	Not applicable / Not sure
1. Understanding what a written chemistry problem is asking you to do	○	○	○	○	○	○
2. Choosing an appropriate equation to solve a chemistry problem	○	○	○	○	○	○
3. Determining appropriate units for a numerical result	○	○	○	○	○	○
4. Reading and writing a chemical formula	○	○	○	○	○	○
5. Describing trends in the periodic table (atomic size, electronegativity)	○	○	○	○	○	○
6. Balancing chemical equations	○	○	○	○	○	○
7. Describing the fundamental structure of an atom	○	○	○	○	○	○
8. Identifying elements that are gases at room temperature (from the periodic table)	○	○	○	○	○	○
9. Converting the temperature in your home from degrees Fahrenheit to kelvin	○	○	○	○	○	○
10.Writing the formula of calcium carbonate, a key ingredient in TUMS	○	○	○	○	○	○
11. Converting your speedometer reading from mph to yards/second (1 mile = 1760 yards)	○	○	○	○	○	○
12. Calculating the density of lemonade (made by adding 50g of lemons to 500mL of water)	○	○	○	○	○	○
13. Identifying the type of change (physical vs. chemical) when milk gets sour	○	○	○	○	○	○
14. Calculating the percent composition of iron in rust (Fe O_2) obtained from your garage door	○	○	○	○	○	○
15. Classifying aluminum foil, salt and salad dressing as compounds, mixtures or elements	○	○	○	○	○	○
16. Explaining why addition of salt melts ice	○	○	○	○	○	○
17. Using chemistry to propose a solution that keeps cooking water from boiling over	○	○	○	○	○	○
18. Writing a summary of the main points of a television documentary that deals with some aspect of chemistry	○	○	○	○	○	○
19. Learning chemistry in this course (if all exams were take-home exams)	○	○	○	○	○	○
20. Doing well on chemistry course exams, given you exert enough effort	○	○	○	○	○	○
21. Asking questions during lecture	○	○	○	○	○	○
22. Learning material in chemistry courses where considerable math is involved	○	○	○	○	○	○
23. Taking an exam or quiz in your chemistry course	○	○	○	○	○	○
24. Taking a chemistry exam or quiz where considerable math is involved	○	○	○	○	○	○
25. Signing up for more chemistry courses in the future (regardless of the outcome of this course or requirements for your major)	○	○	○	○	○	○
26. Preparing for chemistry exams	○	○	○	○	○	○
27. Understanding your chemistry professor	○	○	○	○	○	○
28. Talking to your chemistry professor	○	○	○	○	○	○
29. Receiving the grade you desire in this course	○	○	○	○	○	○
30. Doing homework for this course	○	○	○	○	○	○

APPENDIX E:

Fall 2012 – Survey version with stress and self-efficacy scales

Chemistry Self Efficacy & Anxiety Scale (Fall, 2012)

Name

ID Number

Discussion Section

Please complete the entire survey by filling the bubbles completely (two responses per statement; one for stress and one for confidence). Return this in lecture on Wednesday, December 12th.

Thank you!

	How confident are you about:						How stressful are you about:					
	Not Confident at all	Barely Confident	Neutral	Fairly Confident	Totally Confident	Not applicable / Not sure	Not stressful at all	Barely stressful	Neutral	Fairly stressful	Totally stressful	Not applicable / Not sure
1. Understanding what a written chemistry problem is asking you to do												
2. Choosing an appropriate equation to solve a chemistry problem												
3. Determining appropriate units for a numerical result												
4. Reading and writing a chemical formula												
5. Describing trends in the periodic table (atomic size, electronegativity)												
6. Balancing chemical equations												
7. Describing the fundamental structure of an atom												
8. Identifying elements that are gases at room temperature (from the periodic table)												
9. Converting the temperature in your home from degrees Fahrenheit to kelvin												
10. Writing the formula of calcium carbonate, a key ingredient in TUMS												
11. Converting your speedometer reading from mph to yards/second (1 mile = 1760 yards)												
12. Calculating the density of lemonade (made by adding 50g of lemons to 500mL of water)												

	How confident are you about:						How stressful are you about:					
	Not Confident at all	Barely Confident	Neutral	Fairly Confident	Totally Confident	Not applicable / Not sure	Not stressful at all	Barely stressful	Neutral	Fairly stressful	Totally stressful	Not applicable / Not sure
13. Identifying the type of change (physical vs. chemical) when milk gets sour	○	○	○	○	○	○	○	○	○	○	○	○
14. Calculating the percent composition of iron in rust (Fe O) from your garage door	○	○	○	○	○	○	○	○	○	○	○	○
15. Classifying aluminum foil, salt and salad dressing as compunds, mixtures or elements	○	○	○	○	○	○	○	○	○	○	○	○
16. Explaining why addition of salt melts ice	○	○	○	○	○	○	○	○	○	○	○	○
17. Using chemistry to propose a solution that keeps cooking water from boiling over	○	○	○	○	○	○	○	○	○	○	○	○
18. Writing a summary of the main points of a television documentary that deals with some aspect of chemistry	○	○	○	○	○	○	○	○	○	○	○	○
19. Learning chemistry in this course (if all exams were take-home exams)	○	○	○	○	○	○	○	○	○	○	○	○
20. Doing well on chemistry course exams, given you exert enough effort	○	○	○	○	○	○	○	○	○	○	○	○
21. Asking questions during course lecture	○	○	○	○	○	○	○	○	○	○	○	○
22. Learning material in chemistry courses where considerable math is involved	○	○	○	○	○	○	○	○	○	○	○	○
23. Taking an exam or quiz in your chemistry course	○	○	○	○	○	○	○	○	○	○	○	○
24. Taking a chemistry exam or quiz where considerable math is involved	○	○	○	○	○	○	○	○	○	○	○	○
25. Signing up for more chemistry courses in the future (regardless of the outcome of this course or the requirements for your major)	○	○	○	○	○	○	○	○	○	○	○	○
26. Preparing for chemistry exams	○	○	○	○	○	○	○	○	○	○	○	○
27. Understanding your chemistry professor	○	○	○	○	○	○	○	○	○	○	○	○
28. Talking to your chemistry professor	○	○	○	○	○	○	○	○	○	○	○	○
29. Receiving the grade you desire in this course	○	○	○	○	○	○	○	○	○	○	○	○
30. Doing homework for this course	○	○	○	○	○	○	○	○	○	○	○	○

APPENDIX F:

Fall 2012 – Anxiety survey (standalone)

Chemistry Self Efficacy & Anxiety Survey

ID Number

Name

Please complete the entire survey by filling the bubbles completely (one response per statement) or answering the fill in the blank questions.

Thank you!

How anxious do you get when:

	Not anxious at all	Barely anxious	Neutral	Fairly anxious	Totally anxious	Not applicable / Not sure
1. Signing up for your next chemistry course	O	O	O	O	O	O
2. Getting extra credit for attending your chemistry lecture	O	O	O	O	O	O
3. Learning chemistry in your current and future chemistry courses	O	O	O	O	O	O
4. Hearing the word "chemistry"	O	O	O	O	O	O
5. Learning new concepts in chemistry	O	O	O	O	O	O
6. Walking into your chemistry lecture	O	O	O	O	O	O
7. Taking examinations in your current chemistry course	O	O	O	O	O	O
8. Talking to your chemistry professor	O	O	O	O	O	O
9. Asking or answering questions in your chemistry lecture	O	O	O	O	O	O
10. Cramming the night before your chemistry exam	O	O	O	O	O	O
11. Reading your chemistry textbook to help with homework	O	O	O	O	O	O
12. Listening to lecture in your chemistry class	O	O	O	O	O	O
13. Watching and following your chemistry instructor work a problem on the board	O	O	O	O	O	O
14. Waiting to get a chemistry test returned	O	O	O	O	O	O
15. Not knowing the material on your chemistry exam	O	O	O	O	O	O

How many chemistry courses have you taken, not counting this one?

Have you had laboratory before (chemistry, biology or physics)?

What is your major (intended or declared)?

	very important	important	neither important nor unimportant	unimportant	very unimportant
How important is chemistry in your intended career?	O	O	O	O	O
How important is chemistry in your academic preparation?	O	O	O	O	O

APPENDIX G:

S13 and after – CSEAS online version (Qualtrics)

How confident are you about:

	Not confident at all	Barely confident	Neutral	Fairly confident	Totally confident	Not applicable / Not sure
Understanding what a written chemistry problem is asking you to do	○	○	○	○	○	○
Choosing an appropriate equation to solve a chemistry problem	○	○	○	○	○	○
Determining appropriate units for a numerical result	○	○	○	○	○	○
Reading and writing a chemical formula	○	○	○	○	○	○
Describing trends in the periodic table (atomic size, electronegativity)	○	○	○	○	○	○
Balancing chemical equations	○	○	○	○	○	○
Describing the fundamental structure of an atom	○	○	○	○	○	○
Identifying elements that are gases at room temperature (from the periodic table)	○	○	○	○	○	○
Converting the temperature in your home from degrees Fahrenheit to kelvin	○	○	○	○	○	○
Writing the formula of calcium carbonate, a key ingredient in TUMS	○	○	○	○	○	○

How confident are you about:

	Not confident at all	Barely confident	Neutral	Fairly confident	Totally confident	Not applicable / Not sure
Converting your speedometer reading from mph to yards/second (1 mile = 1760 yards)	○	○	○	○	○	○
Calculating the density of lemonade (made by adding 50 g of lemons to 500 mL of water)	○	○	○	○	○	○
Identifying the type of change (physical vs. chemical) when milk gets sour	○	○	○	○	○	○
Calculating the percent composition of iron in rust (Fe_2O_3) obtained from your garage door	○	○	○	○	○	○
Classifying aluminum foil, salt and salad dressing as compounds, mixtures or elements	○	○	○	○	○	○
Explaining why addition of salt melts ice	○	○	○	○	○	○
Using chemistry to propose a solution that keeps cooking water from boiling over	○	○	○	○	○	○
Writing a summary of the main points of a television documentary that deals with some aspect of chemistry	○	○	○	○	○	○
Learning chemistry in this course (if all exams were take home exams)	○	○	○	○	○	○
Doing well on chemistry course exams, given you exert enough effort	○	○	○	○	○	○

How confident are you about:

	Not confident at all	Barely confident	Neutral	Fairly confident	Totally confident	Not applicable / Not sure
Asking questions during lecture	○	○	○	○	○	○
Learning material in chemistry courses where considerable math is involved	○	○	○	○	○	○
Taking an exam or quiz in your chemistry course	○	○	○	○	○	○
Taking a chemistry exam or quiz where considerable math is involved	○	○	○	○	○	○
Signing up for more chemistry courses in the future (regardless of the outcome of this course or the requirements for your major)	○	○	○	○	○	○
Preparing for chemistry exams	○	○	○	○	○	○
Understanding your chemistry professor	○	○	○	○	○	○
Talking to your chemistry professor	○	○	○	○	○	○
Receiving the grade you desire in this course	○	○	○	○	○	○
Doing homework for this course	○	○	○	○	○	○

How anxious do you get when:

	Not anxious at all	Barely anxious	Neutral	Fairly anxious	Totally anxious	Not applicable / Not sure
Signing up for your next chemistry course						
Getting extra credit for attending your chemistry lecture						
Learning chemistry in your current and future chemistry courses						
Hearing the word "chemistry"						
Learning new concepts in chemistry						
Walking into your chemistry lecture						
Taking examinations in your current chemistry course						
Talking to your chemistry professor						
Asking or answering questions in your chemistry lecture						
Cramming the night before your chemistry exam						
Reading your chemistry textbook to help with homework						
Listening to lecture in your chemistry class						
Watching and following your chemistry instructor work a problem on the board						
Waiting to get a chemistry test returned						
Not knowing the material on your chemistry exam						

Which of the following reasons influenced your decision to choose your current major (intended / declared)? Select all that apply.

If you have a different reason / your major is undecided, please select "Other", but specify your reason/option clearly.

- Inherent interest
- Influence of others
- Materialism
- Good at math/science in high school
- Uninformed choice
- Means to a desired (career) end
- Following family tradition (not pressured)
- Altruism (goodwill / greater good of humanity)
- Negative choice or compromise
- One of several viable options
- Scholarship money available
- Other

How important is chemistry in your intended career?

Very Important	Important	Neither Important nor Unimportant	Unimportant	Very Unimportant

How important is chemistry in your academic preparation?

Very Important	Important	Neither Important nor Unimportant	Unimportant	Very Unimportant

How important are the following factors in determining your confidence level in this course?

	Very Important	Important	Neither Important nor Unimportant	Unimportant	Very Unimportant
Understanding / Learning					
Teaching (by course professor)					
Teaching (by TA/tutor)					
Drive and motivation					
Working with fellow students					
Availability of help (supportive learning culture)					
Grades					
Enjoyment, interest & satisfaction					
Competitive environment					
Means to a desired end (moving on to the next course)					

How important are the following factors in determining your persistence in this course?

	Very Important	Important	Neither Important nor Unimportant	Unimportant	Very Unimportant
Understanding / Learning					
Teaching (by course professor)					
Teaching (by TA/tutor)					
Drive and motivation					
Working with fellow students					
Availability of help (supportive learning culture)					
Grades					
Enjoyment, interest & satisfaction					
Competitive environment					
Means to a desired end (moving on to the next course					

How certain are you of persisting in your current major?

Very Certain	Certain	Neither Certain nor Uncertain	Uncertain	Very Uncertain

Which of the following reasons influenced your response to the previous statement? Select all that apply.

If you have a different reason / your major is undecided, please select "Other", but specify your reason/option clearly.

- ☐ Inherent interest
- ☐ Influence of others
- ☐ Materialism
- ☐ Good at math/science in high school
- ☐ Uninformed choice
- ☐ Means to a desired (career) end
- ☐ Following family tradition (not pressured)
- ☐ Altruism (goodwill / greater good of humanity)
- ☐ Negative choice or compromise
- ☐ One of several viable options
- ☐ Scholarship money available
- ☐ Other (please specify): ______

What is your current major? (If your are intended, please include "intended".)

______

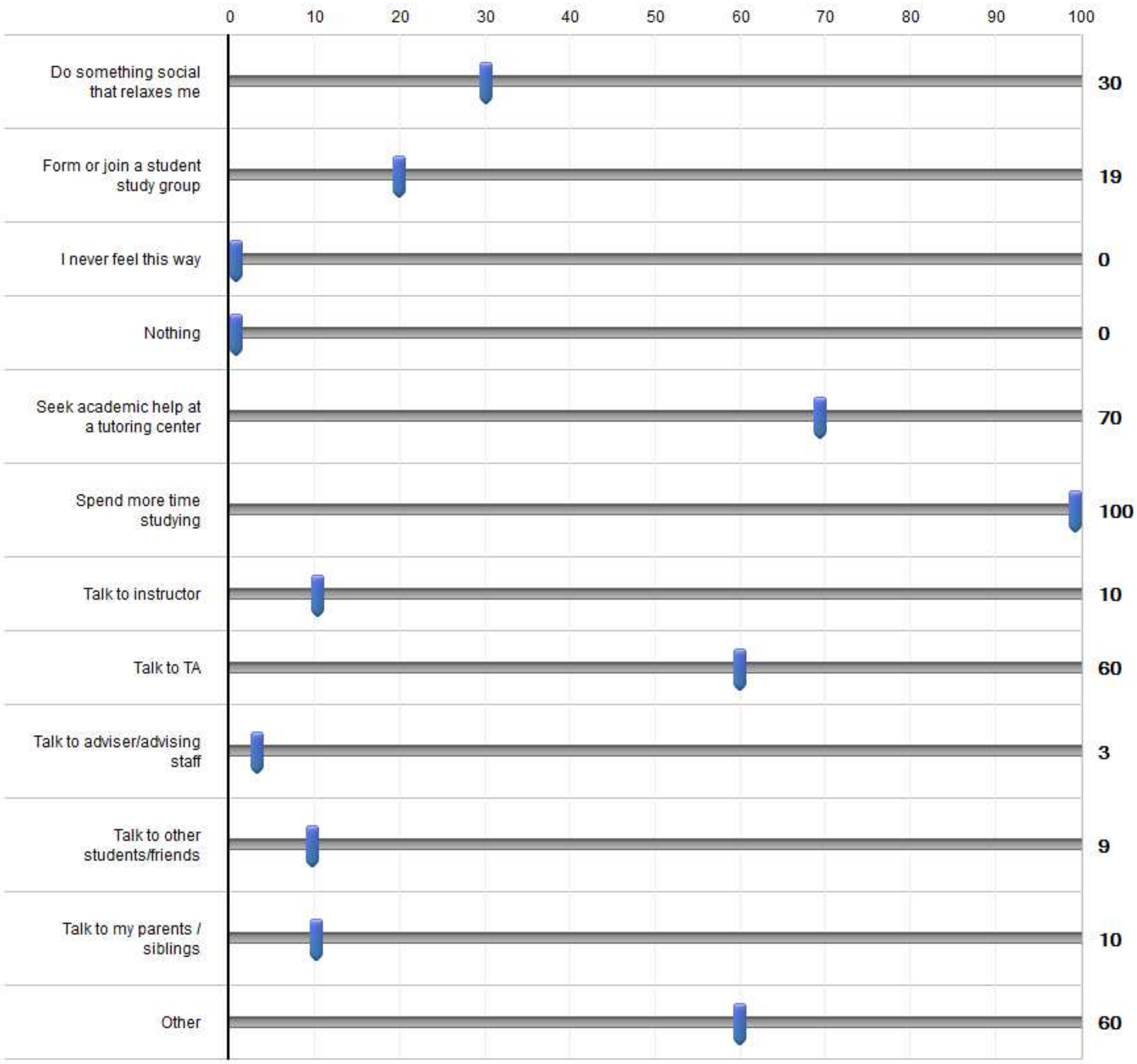
What is your current minor? (If you are intended, please include "intended".)

______

Has your major changed since you began college?

Yes	No	Not applicable
○	○	○

If you struggle in this course, indicate the extent to which you plan to do the following. (Total can be greater than 100)



APPENDIX H:

Results summary – Semantic differential (Spring 2012, GC I)

Data Analysis - Semantic differential (GC I Spring 2012)

Data Collection:

Data were collected using Bauer's Semantic Differential Instrument – a direct method for measuring student attitudes; the instrument consists of a scan sheet with a single word or term at the top of the page and polar adjective pairs or phrases on either side. The word in this study was the single attitude object "chemistry"; adjectives were selected based on how comprehensible they would be to a college-age demographic and also on well they could convey a person's affect regarding chemistry. An in-depth description of the instrument and rationale for adjective choices can be found in (Bauer reference).

Seminal works on semantic differentials originated with Osgood and Tenenbaum, who - through use of factor analysis – isolated three major dimensions of word meanings; the dimensions are evaluation (good or bad), potency (strong or weak) and activity (fast or slow). Adjectives selected for the Differential instrument focus on the evaluation component because this dimension reflects the affective aspect of attitude and typically explains most of the variance.

The analysis described here will focus on pre and post instrument data obtained from students enrolled in a 5-credit introductory college chemistry course during fall 2011 and spring 2012 semesters. The instrument was administered by Teaching Assistants during their respective 1hour discussion sessions with students (week 1 of each semester – pre data). Post data were collected by administering the survey in lecture during the second to last week of each semester. In addition, qualitative data was also obtained in the form of interviews conducted with students taking the same course during summer 2012. Survey responses were manually transcribed to numerical values in the range of 1-7. Statistical tests were performed using Excel and SPSS, in particular.

Results:

Factor Analysis helps reduce data from a group of correlated variables into a smaller set of uncorrelated factors, thus achieving parsimony – explaining the maximum amount of common variance with as few factors as possible.

Exploratory factory analysis was used to identify survey items that show similar response patterns. The following criteria were used to decide if the two-by-two correlation matrix (showing correlations among all the survey items) could be subjected to analysis:

1) Substantial number of correlations in the range of 0.3-0.7.
2) Bartlett's test of sphericity.
3) Measuring of sample adequacy (MSA).
4) Anti-image correlation matrix.
5) Kaiser-Meyer-Olkin measure (KMO).

Data analyzed here satisfied the necessary criteria and were deemed appropriate for factor analysis (distinct factors can be extracted). Presence of 20 variables and at least, if not more than 150 students in each sample also fulfilled the rules for adequate sample size (at least 5 times as many observations as variables). Factors were extracted by the principal components method. Determining the number of factors that could be retained was dictated by a combination of methods:

1) Kaiser's criterion / Eigenvalue > 1
2) Scree plot
3) Fixed percent of variance explained (at least 60-65%).

Since the Eigenvalue condition has been shown to overestimate the number of extracted / retained factors, it was used in conjunction with methods 2 and 3 to determine a reasonable number for retention. 2-7 factors were extracted and the resulting pattern matrix was evaluated for magnitude of loadings, presence of cross loadings and overall structure. Items that loaded with opposite signs were reversed on the scale. Ultimately, the pattern matrix used for comparisons was one that resulted in few cross loadings and struck a balance among percent variance, eigenvalue and scree plot criteria.

Results for pre data obtained in fall 2011 show 4 factors that accounted for 60% of the extracted variance while post data show 3 factors that accounted for 60% of the variance. Attempts to extract more factors resulted in cross loadings and factors which showed only 1 loading. Factors obtained in this study were not given specific names or labels as observed in Bauer's work. Items that constituted each factor did not lean towards one particular affective or semantic category, which made it difficult to collectively summarize each factor with a unique label.

Comparisons between pre and post data for the fall 2011 class suggest that strong item loadings for each factor remained a consistent feature in both data sets. The complete absence of a loading most likely indicates that the item is not conveying anything useful or that the imposed factor structure is in error. From a semantic standpoint, in the fall 2011 pre data, it is also quite possible that a student may not be making an immediate association between the word chemistry and adjective pairs such as scary-fun and insecure-secure, indicating that these items may not be communicating any useful information and thus resulting in no loadings.

While item groupings, in this study, did not follow or come close to those in Bauer's work, Cronbach alpha values for each resulting factor are fairly high, indicating a considerable degree of similarity among the items constituting each factor. This also brings up the possibility that, while responding to an item, a student might go back to an earlier item, similar in meaning, and decide to give an identical response. So, a student could be – intentionally or otherwise – making associations among items.

Factor structures for pre and post data obtained in spring 2012 show 3 factors each with 60% and 59% variance accounted for respectively. Factors display loadings for every item, with the general order of factors staying the same as in fall 2011. Some items undergo reordering while others shift between factors. This "movement" of items once again brings up the question of how students are interpreting chemistry and each adjective pair in the context of chemistry and why items may not be robustly sticking together as more factors are extracted. There is also the possibility that a student could be retaking the course from fall 2011 while a newer student might be taking it as the second course in his or her sequence of introductory chemistry course requirements. In either case, responses to the survey could be impacted by prior chemistry knowledge, experiences, quality of previous instructors, assessments, grades and the overall course structure in general. This might explain why an item that did not have any loadings in fall 2011 data has loadings for spring 2012. Of course, it could very well be the case that a student decided to change his/her mind about a response, did not have an opinion on an item or simply did not put any thought into the survey itself.

Interviews:

In the interests of not inferring reasons without any evidence, interviews were conducted with 2 students during summer 2012 to 1) understand their interpretation of the survey and 2) partly validate possibilities suggested earlier. Participants were 1 male and 1 female (at a large, public, research-oriented Midwestern university) intending to pursue careers in education and geophysics respectively. Male participant had taught 6th grade students but both participants had finished their last chemistry course almost 2 years ago (female participant was retaking the course). Study was approved by the IRB for Human Subjects Protection at the participating academic institution. Participation in the study was entirely voluntary and confirmed with standard informed consent protocols. Data collection consisted of demographic information through use of a short survey and the semantic differential instrument (both of which were completed during the interview).

Basic demographic data were collected through a short survey, including previous science courses taken, career goals and importance of chemistry in fulfilling those goals. Additionally, the participants also completed the semantic differential instrument while rationalizing their responses for each item; this gave the participant and researcher an opportunity to ask/answer questions specifically targeting certain items. The idea of item groupings was never approached by the researcher unless the participant alluded to them during the course of the interview. The interviews were designed to be conversational and were conducted in a safe, non-threatening office used solely for interviews. All interviews were digitally recorded and then transcribed. Notes were taken during the interview to help pace the interviews and to record the researcher's initial reactions.

The interview focused on the participant working through each item in the semantic differential instrument. As a response was chosen, the participant explained the reason for the choice and moved on to the next item. After completing the instrument, demographic questions were answered.

General opinions about the survey

Both students had different opinions on how difficult it was to objectively address the survey without excluding feelings towards instructors (past or present). Male participant, having been a teacher himself, placed a larger burden on the role that instructors play in shaping student attitudes. When talking about this, he said "After all, children first get introduced to chemistry by teachers – whether biological (parents) or academic."

Although the female participant had dropped the course 2 years ago due to problems with an instructor's personality, her inherent interest in Chemistry has sustained her objectivity. So, the instructor is relevant when it comes to her feelings about Chemistry.

Prior experiences shaped a lot of their responses. Male participant drew upon how courses he had taken 2 years ago compared to taking the summer course. Female participant's Montessori pre-school experience sparked her interest in Science and she "went against the grain of society" in public school, where science was "horrible" and math was "hard". According to female participant, the Bauer instrument was open for interpretation but easy to complete and while there was a natural tendency to base responses off of one another, she was able to take the survey objectively.

Male participant, before starting on the first item, asked what "Chemistry is" meant and if it was supposed to be thought of in the realm of the chemistry course, chemistry lab, chemistry as a noun or a whole entity. He also mentioned groupings very early on in the survey and used previous item responses as a guide (responded to confusing/clear based on complicated/simple). He suggested that the survey needed to be clearer in terms of its objective because a lot of adjectives were very vague and it was difficult to answer. Examples of vague adjective pairs, in his opinion, included good/bad (relative to what?) and pleasant/unpleasant.

Specific Interpretations regarding certain adjective pairs:

Both participants thought about safe/dangerous in the context of chemicals – dangerous (nitroglycerin) vs. making rock candy or handling salt. Insecure/Secure was most thought provoking among all the adjective pairs because neither participant believed it had any relation to Chemistry at all. After some serious thought, male participant said, “If we’re talking about Chemistry as a noun, then it is secure. If not, we wouldn’t have the subject matter of Chemistry. There are unknowns in Chemistry, which do make it insecure, but once you debunk theories and gain knowledge, it becomes secure.” Female participant put herself in the context of those two adjectives and said “If I stay on top of my reading, I’m secure. If not, then I’m insecure.” After thinking about an alternate interpretation, she said chemistry is a secure profession and one can pretty much get a job in sciences.

Scary/Fun was perceived differently by both participants as well. While male participant stated that it was scary if studying by himself and would be fun if grades were removed from the equation, the female participant’s interpretation was from an emotive standpoint and whether or not she “got” the material. “If I get it, I have the biggest smile on my face.” Male participant stated that his understanding of specific topics and concepts dictated his response for easy/hard. A topic (subject matter) specific survey resembling the semantic differential would be more meaningful because one needs a situation/environment or a standard basis to make choices. Besides, given that every student comes from different backgrounds and has had varied experiences, a topic specific survey would provide more information. The female participant took a more self-regulated approach and made a choice based on her study habits. Her responses to frustrating/satisfying, complicated/simple and the like were dependent on the effort expended by her and the time she invested in the course.

Summary:

While factor analysis in itself is not a cut and dry statistical method, the instrument is also open to interpretation as observed in the responses given during interviews. 2 students may not form the basis of comprehensive qualitative data, but it can be seen that both students interpreted the survey in different and interesting ways; these varied interpretations might also help understand the lack of a robust or meaningful factor structure.

APPENDIX I:

CSEQ survey

CSXQ College Student Expectations Questionnaire

Welcome!

You have not yet experienced life as a student here. But you have some ideas about how you will spend your time, what you will be doing and so forth. We are interested in these ideas. More specifically, what do you expect to do this year as a student? Please complete the items on the following few pages in a way that answers this question. It takes less than 15 minutes to complete this survey.

Your responses are confidential. Keep in mind that the questionnaire will be read by an electronic scanning device, so be careful in marking your responses. Please use a #2 black lead pencil. Marks made by ink pens cannot be scanned. Do not write or make any marks on the questionnaire outside the spaces for your answers. Erase cleanly any responses you want to change.

The benefits from this or any other survey depend on the thoughtful responses of those who are asked to help. Your willingness to participate is very important and very much appreciated. Thank you!

COLLEGE ACTIVITIES

DIRECTIONS: During the coming year in college, how often do you expect to do the following? Indicate your response by filling in one of the circles to the right of each statement.

Library and Information Technology	Very Often	Often	Occasionally	Never
Use the library as a quiet place to read or study.				
Use an index or database (computer, card catalog, etc.) to find material on some topic.				
Read assigned materials other than textbooks in the library (reserve readings, etc.).				
Develop a bibliography or set of references for a term paper or other report.				
Use a computer or word processor to prepare reports or papers.				
Use e-mail to communicate with an instructor or classmates.				
Participate in class discussions using an electronic medium (e-mail, list-serve, chat group, etc.).				
Search the World Wide Web or Internet for information related to a course.				
Use a computer to retrieve materials from a library not at this institution.				

Experiences with Faculty	Very Often	Often	Occasionally	Never
Ask your instructor for information related to a course you are taking (grades, make-up work, assignments, etc.).				
Discuss your academic program or course selection with a faculty member.				

Experiences with Faculty (cont'd.)	Very Often	Often	Occasionally	Never
Discuss ideas for a term paper or other class project with a faculty member.				
Discuss your career plans and ambitions with a faculty member.				
Socialize with a faculty member outside the classroom (have a snack or soft drink, etc.)				
Ask your instructor for comments and criticisms about your academic performance.				
Work with a faculty member on a research project.				

Course Learning	Very Often	Often	Occasionally	Never
Complete the assigned readings before class.				
Take detailed notes during class.				
Contribute to class discussions.				
Try to see how different facts and ideas fit together.				
Apply material learned in a class to other areas (a job or internship, other courses, relationships with friends, family, co-workers, etc.).				
Summarize major points and information from your readings or class notes.				
Use information or experience from other areas of your life (job, internship, interactions with others) in class discussions or assignments.				
Explain material from a course to someone else (another student, friend, co-worker, family member).				
Prepare a paper or project where you had to integrate ideas from various sources.				

1

APPENDIX J:

Pre-COES (outcome expectations) scale

Introductory Chemistry - Discussion Survey

ID Number

Last Name

First Name

Discussion Section

Instructions: Please fully complete the front and back of this survey. Thank you!

How often do you expect to do the following:

	Never	Occasionally	Often	Very often	Not sure
1. Discuss course information with your instructor (grades, possible make-up work, assignments etc.)	○	○	○	○	○
2. Apply material learned in class to other areas (job / internship, other courses, interactions with others)	○	○	○	○	○
3. Memorize formulas, definitions, technical terms and concepts	○	○	○	○	○
4. Discuss your career plans and ambitions with anyone (for example, advising staff, faculty members, friends or family members)	○	○	○	○	○

Introductory Chemistry - Discussion Survey

Please indicate your level of agreement with each of the statements:

	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
1. If I work hard enough, I will be morely to pass this course.	O	O	O	O	O
2. If I do well/get a good grade in this course, I will be proud of myself.	O	O	O	O	O
3. If I graduate with my current major, I will be more likely to get a well paying job.	O	O	O	O	O
4. If all I do is memorize the solution to any problem solved in lecture/discussion/textbook, I will be successful in this course.	O	O	O	O	O
5. If I change my current major, I will be less likely to get a job.	O	O	O	O	O
6. If I know my interests & abilities, then I will make better career decisions.	O	O	O	O	O
7. If I earn my undergraduate degree, I will be more likely to meet my financial goals.	O	O	O	O	O
8. If I am unable to pass this course, I will be more likely to change my major.	O	O	O	O	O
9. If I try and understand the chemistry while performing an experiment, I will do well in laboratory.	O	O	O	O	O
10. If I make a good career decision, then my family and friends will approve of me.	O	O	O	O	O
11. If I don't understand the concepts in this course, I can pass (with at least a C).	O	O	O	O	O
12. If I obtain a good grade in this course, I will have a better chance of achieving my career goals.	O	O	O	O	O
13. If I can follow the procedure to perform an experiment, I will understand what is happening in laboratory.	O	O	O	O	O
14. If I learn chemistry, I expect to change some of my ideas about how the physical world works.	O	O	O	O	O
15. If I succeed at getting my intended degree, I will be more likely to achieve my career goals.	O	O	O	O	O
16. If I can relate chemistry to situations in my everyday life, I expect to learn it better.	O	O	O	O	O
17. If I figure out what I did wrong on my exam, I will improve my understanding of the material for the next exam.	O	O	O	O	O
18. If I understand a fundamental concept, I can solve homework / exam problems on that concept.	O	O	O	O	O
19. If I can follow my instructor in lecture, I expect to do well in this course.	O	O	O	O	O
20. If I finish my experiment and while in lab, figure out what my data means, I expect to do well in laboratory.	O	O	O	O	O
21. If I can explain a problem or concept to a classmate, I will understand the material better.	O	O	O	O	O
22. If I do everything possible (for example, review class notes, read the textbook, solve several sample problems, do homework, maintain perfect attendance), I will do well in this course.	O	O	O	O	O
23. If I understand the principles behind the experiments, I will be more likely to succeed in laboratory.	O	O	O	O	O
24. If I do everything possible (for example, review class notes, read the textbook, solve several sample problems, do homework, maintain perfect attendance), I will be prepared for quizzes / exams in this course.	O	O	O	O	O
25. If I can remember the solution to a problem and know where to put the numbers, I will do well on quizzes / exams in this course.	O	O	O	O	O

APPENDIX K:

Post-COES (outcome expectations) scale

Introductory Chemistry - End of Semester Survey

ID Number

Last Name

First Name

Discussion Section

Instructions: Please fully complete the front and back of this survey. Thank you!

How often did you do the following:

	Never	Occasionally	Often	Very often	Not sure
1. Discuss course information with your instructor (grades, possible make-up work, assignments etc.)	O	O	O	O	O
2. Apply material learned in class to other areas (job / internship, other courses, interactions with others)	O	O	O	O	O
3. Memorize formulas, definitions, technical terms and concepts	O	O	O	O	O
4. Discuss your career plans and ambitions with anyone (for example, advising staff, faculty members, friends or family members)	O	O	O	O	O

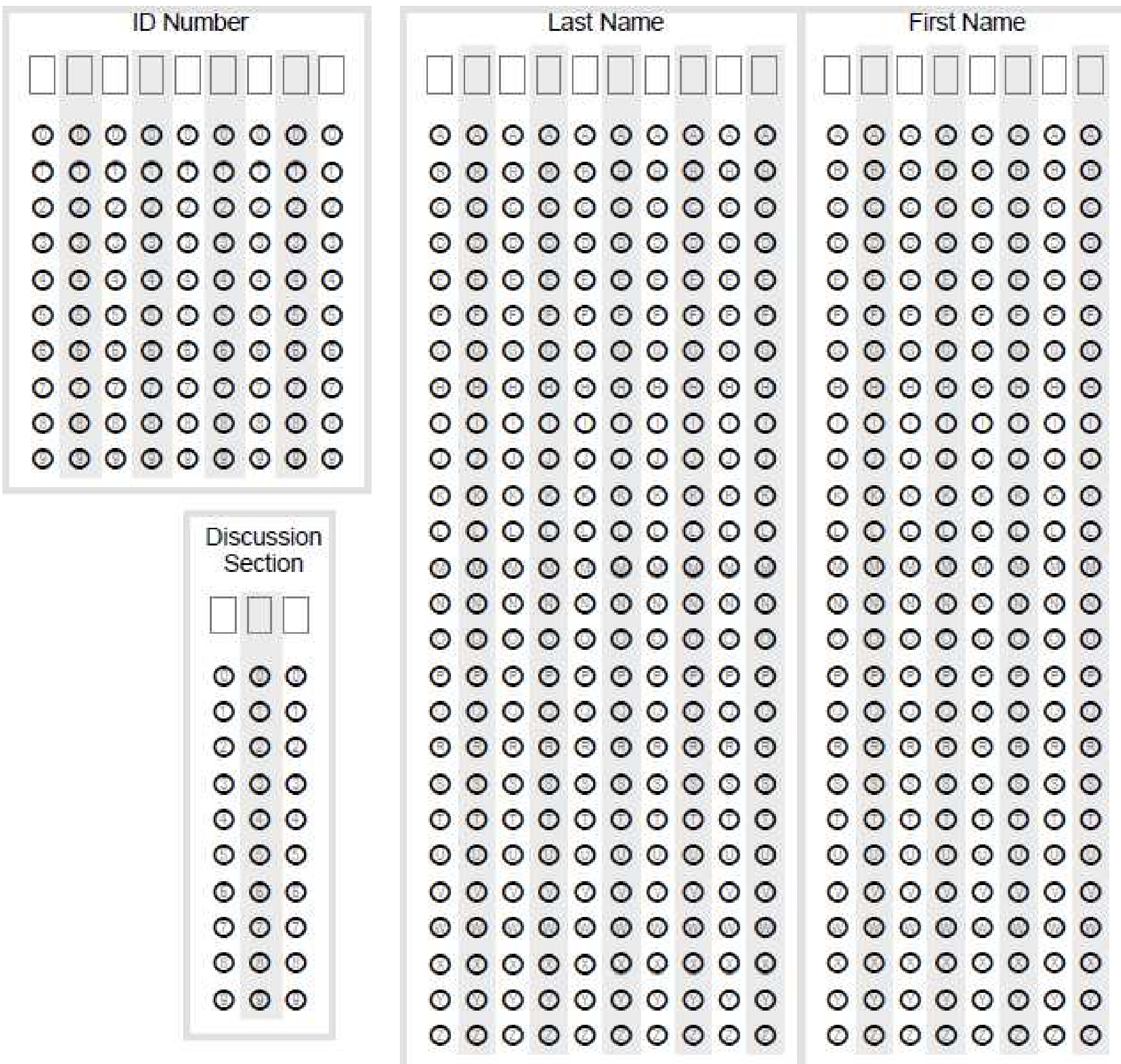
Introductory Chemistry - End of Semester Survey

Please indicate your level of agreement with each of the statements:	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
1. If I work hard enough, I will be more likely to pass a future.	○	○	○	○	○
2. If I do well/get a good grade in this course, I will be proud of myself.	○	○	○	○	○
3. If I graduate with my current major, I will be more likely to get a well paying job.	○	○	○	○	○
4. If all I did was memorize the solution to any problem solved in lecture/discussion/textbook, I can be successful in a future chemistry course.	○	○	○	○	○
5. If I change my current major, I will be less likely to get a job.	○	○	○	○	○
6. If I know my interests & abilities, then I will make better career decisions.	○	○	○	○	○
7. If I earn my undergraduate degree, I will be able to meet my financial goals.	○	○	○	○	○
8. If I am unable to pass this course, I will be more likely to change my major.	○	○	○	○	○
9. If I try and understand the chemistry while performing an experiment, I can do better in a future laboratory course.	○	○	○	○	○
10. If I make a good career decision, then my family and friends will approve of me.	○	○	○	○	○
11. If I don't understand the concepts in this course, I can pass (with at least a C).	○	○	○	○	○
12. If I obtain a good grade in this course, I will have a better chance of achieving my career goals.	○	○	○	○	○
13. If I am able to follow the procedure to perform an experiment, I can understand what is happening in a future chemistry laboratory course.	○	○	○	○	○
14. If I learn chemistry, I expect to change some of my ideas about how the physical world works.	○	○	○	○	○
15. If I succeed at getting my intended degree, I will be more likely to achieve my career goals.	○	○	○	○	○
16. If I can relate chemistry to situations in my everyday life, I expect to learn it better.	○	○	○	○	○
17. If I figure out what I did wrong on my exam, I will improve my understanding of course material for the next exam.	○	○	○	○	○
18. If I understand a fundamental concept, I can solve homework / exam problems on that concept.	○	○	○	○	○
19. If I can follow my instructor in lecture, I expect to do better in a future chemistry course.	○	○	○	○	○
20. If I finish my experiment and while in lab, figure out what my data means, I can do well in a future chemistry laboratory course.	○	○	○	○	○
21. If I can explain a problem or concept to a classmate, I will understand the material better.	○	○	○	○	○
22. If I do everything possible (for example, review class notes, read the textbook, solve several sample problems, do homework, maintain perfect attendance), I can do well in a future chemistry course.	○	○	○	○	○
23. If I understand the principles behind the experiments, I will be more likely to succeed in a future chemistry laboratory course.	○	○	○	○	○
24. If I do everything possible (for example, review class notes, read the textbook, solve several sample problems, do homework, maintain perfect attendance), I can be prepared for quizzes / exams in a future chemistry course.	○	○	○	○	○
25. If I can remember the solution to a problem and know where to put the numbers, I can do well on quizzes / exams in a future chemistry course.	○	○	○	○	○

APPENDIX L:

Subset (shortened) instrument

Introductory Chemistry - Lecture Survey, Pre-Exam 1



Instructions: Please fully complete the front and back of this survey.

Return this in lecture on Wednesday, September 29, 2014.

For completing this and the post exam 1 survey (handed out during the week of October 6th), you will receive 2 extra credit points on exam 1.

Thank you!

Introductory Chemistry - Lecture Survey

How confident are you about:	Not Confident at all	Barely Confident	Neutral	Fairly Confident	Totally Confident	Not applicable / not sure
1. Choosing an appropriate equation to solve a chemistry problem.	O	O	O	O	O	O
2. Determining appropriate units for a numerical result.	O	O	O	O	O	O
3. Reading and writing a chemical formula.	O	O	O	O	O	O
4. Describing trends in the periodic table (atomic size, electronegativity).	O	O	O	O	O	O
5. Identifying the type of change (physical vs. chemical) when milk gets sour.	O	O	O	O	O	O
6. Converting your speedometer reading from mph to yards/sec (1 mile = 1760 yards).	O	O	O	O	O	O
7. Calculating the density of lemonade (made by adding 50 g lemons to 500 mL of water).	O	O	O	O	O	O
8. Calculating the percent composition of iron in rust (Fe2O3) obtained from your garage door.	O	O	O	O	O	O
9. Taking an exam or quiz in your chemistry course.	O	O	O	O	O	O
10. Preparing for chemistry exams.	O	O	O	O	O	O
11. Understanding your chemistry professor.	O	O	O	O	O	O
12. Talking to your chemistry professor.	O	O	O	O	O	O
13. Receiving the grade you desire in this course.	O	O	O	O	O	O

Please indicate your level of agreement with each of the statements:	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
14. If all I do is memorize the solution to any problem solved in lecture/discussion/textbook, I will be successful in this course.	O	O	O	O	O
15. If I know my interests & abilities, then I will make better career decisions.	O	O	O	O	O
16. If I try and understand the chemistry while performing an experiment, I will do well in laboratory.	O	O	O	O	O
17. If I don't understand the concepts in this course, I can pass (with at least a C).	O	O	O	O	O
18. If I obtain a good grade in this course, I will have a better chance of achieving my career goals.	O	O	O	O	O
19. If I learn chemistry, I expect to change some of my ideas about how the physical world works.	O	O	O	O	O
20. If I succeed at getting my intended degree, I will be more likely to achieve my career goals.	O	O	O	O	O
21. If I can relate chemistry to situations in my everyday life, I expect to learn it better.	O	O	O	O	O
22. If I understand a fundamental concept, I can solve homework / exam problems on that concept.	O	O	O	O	O
23. If I finish my experiment and while in lab, figure out what my data means, I expect to do well in laboratory.	O	O	O	O	O
24. If I do everything possible (for example, review class notes, read the textbook, solve several sample problems, do homework, maintain perfect attendance), I will do well in this course.	O	O	O	O	O
25. If I do everything possible (for example, review class notes, read the textbook, solve several sample problems, do homework, maintain perfect attendance), I will be prepared for quizzes / exams in this course.	O	O	O	O	O

APPENDIX M:

STEM major designations – as of May 2016

STEM Designated Degree Program List Effective May 10, 2016

The STEM Designated Degree Program list is a complete list of fields of study that DHS considers to be science, technology, engineering or mathematics (STEM) fields of study for purposes of the 24-month STEM optional practical training extension described at 8 CFR 214.2(f). Under 8 CFR 214.2(f)(10)(ii)(C)(2), a STEM field of study is a field of study "included in the Department of Education's Classification of Instructional Programs taxonomy within the two-digit series containing engineering, biological sciences, mathematics, and physical sciences, or a related field. In general, related fields will include fields involving research, innovation, or development of new technologies using engineering, mathematics, computer science, or natural sciences (including physical, biological, and agricultural sciences)."

Accordingly, this list designates the following four CIP summary groups/series at the 2-digit CIP code level: Engineering (CIP code 14), Biological and Biomedical Sciences (CIP code 26), Mathematics and Statistics (CIP code 27), and Physical Sciences (CIP code 40). Any new additions to those areas will automatically be included on this STEM Designated Degree Program list. Consistent with the definition of "related field" above, related fields in this list include fields involving research, innovation, or development of new technologies using engineering, mathematics, computer science, or natural sciences. DHS designates these fields at the 6-digit level.

CIP Code Two-Digit Series	2010 CIP Code	CIP Code Title
01	01.0308	Agroecology and Sustainable Agriculture
01	01.0901	Animal Sciences, General
01	01.0902	Agricultural Animal Breeding
01	01.0903	Animal Health
01	01.0904	Animal Nutrition
01	01.0905	Dairy Science
01	01.0906	Livestock Management
01	01.0907	Poultry Science
01	01.0999	Animal Sciences, Other
01	01.1001	Food Science
01	01.1002	Food Technology and Processing
01	01.1099	Food Science and Technology, Other
01	01.1101	Plant Sciences, General
01	01.1102	Agronomy and Crop Science
01	01.1103	Horticultural Science
01	01.1104	Agricultural and Horticultural Plant Breeding
01	01.1105	Plant Protection and Integrated Pest Management
01	01.1106	Range Science and Management
01	01.1199	Plant Sciences, Other
01	01.1201	Soil Science and Agronomy, General
01	01.1202	Soil Chemistry and Physics
01	01.1203	Soil Microbiology
01	01.1299	Soil Sciences, Other
03	03.0101	Natural Resources/Conservation, General
03	03.0103	Environmental Studies
03	03.0104	Environmental Science
03	03.0199	Natural Resources Conservation and Research, Other
03	03.0205	Water, Wetlands, and Marine Resources Management

15	15.0614	Welding Engineering Technology/Technician
15	15.0615	Chemical Engineering Technology/Technician
15	15.0616	Semiconductor Manufacturing Technology
15	15.0699	Industrial Production Technologies/Technicians, Other
15	15.0701	Occupational Safety and Health Technology/Technician
15	15.0702	Quality Control Technology/Technician
15	15.0703	Industrial Safety Technology/Technician
15	15.0704	Hazardous Materials Information Systems Technology/Technician
15	15.0799	Quality Control and Safety Technologies/Technicians, Other
15	15.0801	Aeronautical/Aerospace Engineering Technology/Technician
15	15.0803	Automotive Engineering Technology/Technician
15	15.0805	Mechanical Engineering/Mechanical Technology/Technician
15	15.0899	Mechanical Engineering Related Technologies/Technicians, Other
15	15.0901	Mining Technology/Technician
15	15.0903	Petroleum Technology/Technician
15	15.0999	Mining and Petroleum Technologies/Technicians, Other
15	15.1001	Construction Engineering Technology/Technician
15	15.1102	Surveying Technology/Surveying
15	15.1103	Hydraulics and Fluid Power Technology/Technician
15	15.1199	Engineering-Related Technologies, Other
15	15.1201	Computer Engineering Technology/Technician
15	15.1202	Computer Technology/Computer Systems Technology
15	15.1203	Computer Hardware Technology/Technician
15	15.1204	Computer Software Technology/Technician
15	15.1299	Computer Engineering Technologies/Technicians, Other
15	15.1301	Drafting and Design Technology/Technician, General
15	15.1302	CAD/CADD Drafting and/or Design Technology/Technician
15	15.1303	Architectural Drafting and Architectural CAD/CADD
15	15.1304	Civil Drafting and Civil Engineering CAD/CADD
15	15.1305	Electrical/Electronics Drafting and Electrical/Electronics CAD/CADD
15	15.1306	Mechanical Drafting and Mechanical Drafting CAD/CADD
15	15.1399	Drafting/Design Engineering Technologies/Technicians, Other
15	15.1401	Nuclear Engineering Technology/Technician
15	15.1501	Engineering/Industrial Management
15	15.1502	Engineering Design
15	15.1503	Packaging Science
15	15.1599	Engineering-Related Fields, Other
15	15.1601	Nanotechnology
15	15.9999	Engineering Technologies and Engineering-Related Fields, Other
26	26.XXXX	Biological and Biomedical Sciences
27	27.XXXX	Mathematics and Statistics
28	28.0501	Air Science/Airpower Studies
28	28.0502	Air and Space Operational Art and Science
28	28.0505	Naval Science and Operational Studies
29	29.0201	Intelligence, General
29	29.0202	Strategic Intelligence
29	29.0203	Signal/Geospatial Intelligence
29	29.0204	Command & Control (C3, C4I) Systems and Operations
29	29.0205	Information Operations/Joint Information Operations
29	29.0206	Information/Psychological Warfare and Military Media Relations
29	29.0207	Cyber/Electronic Operations and Warfare
29	29.0299	Intelligence, Command Control and Information Operations, Other
29	29.0301	Combat Systems Engineering
29	29.0302	Directed Energy Systems
29	29.0303	Engineering Acoustics
29	29.0304	Low-Observables and Stealth Technology
29	29.0305	Space Systems Operations
29	29.0306	Operational Oceanography
29	29.0307	Undersea Warfare
29	29.0399	Military Applied Sciences, Other
29	29.0401	Aerospace Ground Equipment Technology
29	29.0402	Air and Space Operations Technology
29	29.0403	Aircraft Armament Systems Technology
29	29.0404	Explosive Ordinance/Bomb Disposal
29	29.0405	Joint Command/Task Force (C3, C4I) Systems

03	03.0502	Forest Sciences and Biology
03	03.0508	Urban Forestry
03	03.0509	Wood Science and Wood Products/Pulp and Paper Technology
03	03.0601	Wildlife, Fish and Wildlands Science and Management
04	04.0902	Architectural and Building Sciences/Technology
09	09.0702	Digital Communication and Media/Multimedia
10	10.0304	Animation, Interactive Technology, Video Graphics and Special Effects
11	11.0101	Computer and Information Sciences, General
11	11.0102	Artificial Intelligence
11	11.0103	Information Technology
11	11.0104	Informatics
11	11.0199	Computer and Information Sciences, Other
11	11.0201	Computer Programming/Programmer, General
11	11.0202	Computer Programming, Specific Applications
11	11.0203	Computer Programming, Vendor/Product Certification
11	11.0299	Computer Programming, Other
11	11.0301	Data Processing and Data Processing Technology/Technician
11	11.0401	Information Science/Studies
11	11.0501	Computer Systems Analysis/Analyst
11	11.0701	Computer Science
11	11.0801	Web Page, Digital/Multimedia and Information Resources Design
11	11.0802	Data Modeling/Warehousing and Database Administration
11	11.0803	Computer Graphics
11	11.0804	Modeling, Virtual Environments and Simulation
11	11.0899	Computer Software and Media Applications, Other
11	11.0901	Computer Systems Networking and Telecommunications
11	11.1001	Network and System Administration/Administrator
11	11.1002	System, Networking, and LAN/WAN Management/Manager
11	11.1003	Computer and Information Systems Security/Information Assurance
11	11.1004	Web/Multimedia Management and Webmaster
11	11.1005	Information Technology Project Management
11	11.1006	Computer Support Specialist
11	11.1099	Computer/Information Technology Services Administration and Management, Other
13	13.0501	Educational/Instructional Technology
13	13.0601	Educational Evaluation and Research
13	13.0603	Educational Statistics and Research Methods
14	14.XXXX	Engineering
15	15.0000	Engineering Technology, General
15	15.0101	Architectural Engineering Technology/Technician
15	15.0201	Civil Engineering Technology/Technician
15	15.0303	Electrical, Electronic and Communications Engineering Technology/Technician
15	15.0304	Laser and Optical Technology/Technician
15	15.0305	Telecommunications Technology/Technician
15	15.0306	Integrated Circuit Design
15	15.0399	Electrical and Electronic Engineering Technologies/Technicians, Other
15	15.0401	Biomedical Technology/Technician
15	15.0403	Electromechanical Technology/Electromechanical Engineering Technology
15	15.0404	Instrumentation Technology/Technician
15	15.0405	Robotics Technology/Technician
15	15.0406	Automation Engineer Technology/Technician
15	15.0499	Electromechanical and Instrumentation and Maintenance Technologies/Technicians, Other
15	15.0501	Heating, Ventilation, Air Conditioning and Refrigeration Engineering Technology/Technician
15	15.0503	Energy Management and Systems Technology/Technician
15	15.0505	Solar Energy Technology/Technician
15	15.0506	Water Quality and Wastewater Treatment Management and Recycling Technology/Technician
15	15.0507	Environmental Engineering Technology/Environmental Technology
15	15.0508	Hazardous Materials Management and Waste Technology/Technician
15	15.0599	Environmental Control Technologies/Technicians, Other
15	15.0607	Plastics and Polymer Engineering Technology/Technician
15	15.0611	Metallurgical Technology/Technician
15	15.0612	Industrial Technology/Technician
15	15.0613	Manufacturing Engineering Technology/Technician

15	15.0614	Welding Engineering Technology/Technician
15	15.0615	Chemical Engineering Technology/Technician
15	15.0616	Semiconductor Manufacturing Technology
15	15.0699	Industrial Production Technologies/Technicians, Other
15	15.0701	Occupational Safety and Health Technology/Technician
15	15.0702	Quality Control Technology/Technician
15	15.0703	Industrial Safety Technology/Technician
15	15.0704	Hazardous Materials Information Systems Technology/Technician
15	15.0799	Quality Control and Safety Technologies/Technicians, Other
15	15.0801	Aeronautical/Aerospace Engineering Technology/Technician
15	15.0803	Automotive Engineering Technology/Technician
15	15.0805	Mechanical Engineering/Mechanical Technology/Technician
15	15.0899	Mechanical Engineering Related Technologies/Technicians, Other
15	15.0901	Mining Technology/Technician
15	15.0903	Petroleum Technology/Technician
15	15.0999	Mining and Petroleum Technologies/Technicians, Other
15	15.1001	Construction Engineering Technology/Technician
15	15.1102	Surveying Technology/Surveying
15	15.1103	Hydraulics and Fluid Power Technology/Technician
15	15.1199	Engineering-Related Technologies, Other
15	15.1201	Computer Engineering Technology/Technician
15	15.1202	Computer Technology/Computer Systems Technology
15	15.1203	Computer Hardware Technology/Technician
15	15.1204	Computer Software Technology/Technician
15	15.1299	Computer Engineering Technologies/Technicians, Other
15	15.1301	Drafting and Design Technology/Technician, General
15	15.1302	CAD/CADD Drafting and/or Design Technology/Technician
15	15.1303	Architectural Drafting and Architectural CAD/CADD
15	15.1304	Civil Drafting and Civil Engineering CAD/CADD
15	15.1305	Electrical/Electronics Drafting and Electrical/Electronics CAD/CADD
15	15.1306	Mechanical Drafting and Mechanical Drafting CAD/CADD
15	15.1399	Drafting/Design Engineering Technologies/Technicians, Other
15	15.1401	Nuclear Engineering Technology/Technician
15	15.1501	Engineering/Industrial Management

15	15.1502	Engineering Design
15	15.1503	Packaging Science
15	15.1599	Engineering-Related Fields, Other
15	15.1601	Nanotechnology
15	15.9999	Engineering Technologies and Engineering-Related Fields, Other
26	26.XXXX	Biological and Biomedical Sciences
27	27.XXXX	Mathematics and Statistics
28	28.0501	Air Science/Airpower Studies
28	28.0502	Air and Space Operational Art and Science
28	28.0505	Naval Science and Operational Studies
29	29.0201	Intelligence, General
29	29.0202	Strategic Intelligence
29	29.0203	Signal/Geospatial Intelligence
29	29.0204	Command & Control (C3, C4I) Systems and Operations
29	29.0205	Information Operations/Joint Information Operations
29	29.0206	Information/Psychological Warfare and Military Media Relations
29	29.0207	Cyber/Electronic Operations and Warfare
29	29.0299	Intelligence, Command Control and Information Operations, Other
29	29.0301	Combat Systems Engineering
29	29.0302	Directed Energy Systems
29	29.0303	Engineering Acoustics
29	29.0304	Low-Observables and Stealth Technology
29	29.0305	Space Systems Operations
29	29.0306	Operational Oceanography
29	29.0307	Undersea Warfare
29	29.0399	Military Applied Sciences, Other
29	29.0401	Aerospace Ground Equipment Technology
29	29.0402	Air and Space Operations Technology
29	29.0403	Aircraft Armament Systems Technology
29	29.0404	Explosive Ordinance/Bomb Disposal
29	29.0405	Joint Command/Task Force (C3, C4I) Systems

APPENDIX N:

Results of one-way ANOVA to select predictors – persistence model

One-way ANOVA to determine significant predictors for logistic regression - persistence model

		Sum of squares	df	Mean Square	F	Sig.
Self-efficacy related to applying chemistry strategies	Between Groups	4.076	3	1.359	3.125	**0.026**
	Within Groups	123.924	285	0.435		
	Total	128.000	288			

CURRICULUM VITAE

EDUCATION

Ph.D. Chemistry, Expected May 2017
University of Wisconsin – Milwaukee, WI
Dissertation: *Development and testing of a longitudinal STEM persistence model integrating self-efficacy, outcome expectations and performance.*

M.S. Forensic Science, Dec 2007
University of Illinois – Chicago, IL
Project: *Creating a reference library of hair and fiber standards using FT-IR/ATR*.

M.S. Inorganic Chemistry, Dec 2005
Purdue University – West Lafayette, IN
Thesis: *Alloy mediated synthesis of metal-semiconductor composites.*

B.S. Chemistry, Dec 2001
Purdue University – West Lafayette, IN
Project: *Examining metal-mediated protein cross links in a mussel adhesive.*

RESEARCH EXPERIENCE

Doctoral research - **Chemical Education** Jan 2012 to May 2017
University of Wisconsin – Milwaukee, WI
Advisor: Dr. Kristen Murphy

- Developed and adapted surveys to assess students' outcome expectations and self-efficacy in several general chemistry courses
- Established interview protocols, conducted student interviews and analyzed resulting data to validate surveys and develop affective profiles for at-risk students
- Designed and tested interventions (using Qualtrics) to offset lack of persistence in at-risk student groups
- Developed and tested eye-tracking stimuli to assess usability and efficacy of interventions

Research Assistant (Summer) – **Chemical Education** Jan 2013 to Dec 2016
University of Wisconsin – River Falls (UWRF) and University of Wisconsin – Milwaukee
PI: Dr. Jamie Schneider and Co-PI: Dr. Kristen Murphy
The effect of feedback on multiple - choice chemistry assessment (***NSF - DUE 1140914***)

- Evaluated test statistics for chemistry assessments (practice exams) offered at UWRF
- Examined impact of various feedback mechanisms on exam performance
- Assessed exam complexity (using complexity ratings from multiple researchers), student confidence and their combined effects on exam performance under varied feedback modes
- Collaborated with postdoctoral researcher from UWRF to conduct analyses and present results during weekly meetings

Research Assistant – Nanomaterials Chemistry Jan 2009 to June 2009
Florida International University, Miami, FL

- Synthesized water-soluble conjugated polymer nanoparticles (CPNs), via organic routes, for use in drug testing
- Conducted size and molecular weight analyses of CPNs using SEM and ZetaSizer Nano respectively
- Managed and supervised laboratory inventory, safety protocol and general maintenance of the research laboratory

Research Assistant – Materials Chemistry June 2002 to Dec 2005
Purdue University – West Lafayette, IN

- Synthesized metal-semiconductor composite electrodes via an inter-metallic alloy mediated synthetic route
- Examined properties of electrodes using various techniques –microscopy (SEM/EDS, TEM, AFM), and X-ray diffraction
- Supervised research projects of two undergraduate students and trained them on the use of laboratory equipment and techniques.

SELECTED RESEARCH SKILLS / TECHNIQUES

- Managing, organizing and formatting large, complex datasets
- Employing appropriate methods to explore and analyze qualitative and quantitative data
- Expertise in a wide range of multivariate statistical methods **– factor analyses (exploratory and confirmatory), cluster analyses, multiple and logistic regression, discriminant analyses, Monte carlo PCA for parallel analysis**
- Developing and evaluating psychometric properties of assessments and surveys
- Establishing protocols and stimuli for conducting student interviews – **response process, eye tracking, transcribing and coding**
- Proficiency in SPSS, SAS, JMP, Excel, Qualtrics

CONFERENCE PRESENTATIONS

- Srinivasan, S.[*]; Murphy, K.L. (2016). "Development and preliminary testing of a STEM persistence model", *24th Biennial Conference on Chemical Education*, University of Northern Colorado, Greeley, CO. (**Paper)**

- Srinivasan, S.[*]; Murphy, K.L. (2016). "Affective profiles to target at-risk students: Evaluating the impact of interventions on students' STEM persistence", *24th Biennial Conference on Chemical Education*, University of Northern Colorado, Greeley, CO. (**Paper**)

- Srinivasan, S.[*]; Murphy, K.L. (2016). "Development and preliminary testing of a STEM persistence model: Using a subset instrument to generate affective profiles", *251st ACS National Meeting*, San Diego, CA. **(Paper**)

- Srinivasan, S.[*]; Murphy, K.L. (2015). "Persistence in STEM: Using a subset instrument to measure subtle changes in self-efficacy and outcome expectations*"*, *250th ACS National Meeting*, Boston, MA. **(Paper**)

- Srinivasan, S.[*]; Murphy, K.L. (2014). "Development and preliminary testing of a persistence instrument: Measuring outcome expectations", *23rd Biennial Conference on Chemical Education*, Grand Valley State University, Allendale, MI. (**Paper)**

- Srinivasan, S.[*]; Murphy, K.L. (2014). "Development and preliminary testing of a persistence instrument: Measuring self-efficacy and outcome expectations", *Engendering Change: The 4th Annual Gender and Sexualities Graduate Student Conference*, Northwestern University, Evanston, IL. (**Paper**)

- Srinivasan, S.[*]; Murphy, K.L. (2013). "Development and preliminary testing of a persistence instrument: Measuring self-efficacy", *Chemistry Education Research Conference,* Miami University, Oxford, OH. (**Poster)**

- Srinivasan, S.[*]; Murphy, K.L. (2013). "Development and preliminary testing of a persistence instrument: Measuring self-efficacy", *Sci-Mix: 244th ACS National Meeting*, Indianapolis, Indiana. **(Poster)**

- Srinivasan, S.[*]; Murphy, K.L. (2013). "Development and preliminary testing of a persistence instrument: Measuring self-efficacy", *244th ACS National Meeting*, Indianapolis, IN. (**Paper**)

- Srinivasan, S.[*]; Murphy, K.L. (2013). "Development and preliminary testing of a persistence instrument: Measuring self-efficacy", *40th Annual Great Lakes Regional Meeting*, La Crosse, WI. (**Paper**)

DEPARTMENTAL PRESENTATIONS

- Srinivasan, S.*; Murphy, K.L. (2015). "Persistence in STEM: Using a subset instrument to measures subtle changes in self-efficacy and outcome expectations", *Department of Chemistry and Biochemistry Awards Day Symposium*, University of Wisconsin, Milwaukee, WI. (**Poster**)

- Srinivasan, S.*; Murphy, K.L. (2014). "Development and preliminary testing of a persistence instrument: Measuring outcome expectations", *Department of Chemistry and Biochemistry Awards Day Symposium*, University of Wisconsin, Milwaukee, WI. (**Poster**)

- Srinivasan, S.*; Murphy, K.L. (2014). "Development and preliminary testing of a persistence instrument: Measuring self-efficacy", *Department of Chemistry and Biochemistry Awards Day Symposium*, University of Wisconsin, Milwaukee, WI. (**Poster**)

MANUSCRIPTS IN PREPARATION

- Srinivasan S., Murphy K. (2017). *Development and preliminary testing of a persistence instrument: Measuring outcome expectations.*

- Srinivasan S., Murphy K. (2017). *Development and preliminary testing of a persistence instrument: Measuring self-efficacy.*

- Srinivasan S., Murphy K. (2017). *Using a shortened subset instrument to develop affective profiles: Tracking finer changes in the persistence model*

- Srinivasan S., Murphy K. (2017). *Integrating performance and affective measures in the development of a longitudinal STEM persistence model*

PUBLISHED PAPERS

- Tan Y, Srinivasan S, Choi K.S. (2005). Electrochemical deposition of mesoporous nickel hydroxide films from dilute surfactant solutions**.** *Journal of the American Chemical Society*, 127, 3596-3604.

- Sever M.J., Weisser J.T., Monahan J, Srinivasan S, Wilker J.J. (2004). Metal Mediated cross-linking in the generation of a marine mussel adhesive**.** *Angewandte Chemie - International Edition*, 43, 448-450.

SERVICE

Profession – Meeting organization:

- ***Co-organizer and Presider*** of a symposium for the *23rd Biennial Conference on Chemical Education*, Grand Valley State University, Allendale, MI
  Importance of the affective domain in research and teaching (2014); Shalini Srinivasan and Kristen Murphy (University of Wisconsin – Milwaukee)

- ***Chair and Organizer*** of a poster social & mixer for the *Milwaukee section of the American Chemical Society's Younger Chemists Committee*, University of Wisconsin, Milwaukee, WI (2013)

Profession – Membership:

- Member, American Chemical Society. Jan 2012 – present

HONORS AND AWARDS

- Chancellor's Graduate Fellowship Jan 2012 – present
- Teaching assistant of the year – General chemistry for engineers 2014 – 2015
- 3rd place best poster – Chemistry department research symposium April 2015
- Chemistry Department Graduate Student Travel Award May 2015
- Mentoring Travel Award – New Graduate Student Mentor Program 2015 – 2016
- Dean's List & Semester honors Aug. 2000 to Dec 2001

CURRICULUM DEVELOPMENT

Chemistry for Engineers (CHEM 105) Jan 2015 – May 2016

- Coordinated and led teaching efforts in support of "**flipped classroom**" model for a class of 150 students – **sole** teaching assistant for five discussion sections
- Designed worksheets and exam review packets for class discussions and exams respectively; elicited input and coordinated alignment with course faculty/students.
- Initiated class evaluation process for continuous feedback and improvement.
- Led weekly supplemental instruction (**SI leader**) sessions for students
  - Integrated course content with study strategies based on student progress reports from ALEKS
  - Discussed key concepts and helped students organize class material
  - Evaluated success of each lesson module in preparation for next SI session

MENTORING

Graduate students

- Past and current mentor for incoming chemistry graduate teaching assistants as part of the mentoring program at University of Wisconsin – Milwaukee
- Evaluated designated mentees during their teaching sessions and provided reports to department co-chair
- Assisted mentees with preparing course materials – discussion worksheets and quizzes.

Undergraduate students

- Mentored undergraduate students in several chemical education research projects
- Guided students in formulating a research hypothesis and developing proposals
- Assisted students in preparing posters to present at departmental research symposia

    - "The use of reported confidence versus reported mental workload of students in introductory chemistry when working informal review items" – Nicole Endres (5th place best undergraduate poster at the Chemistry department research symposium)

    - "Student attitudes towards the subject of chemistry inventory: Changes during a first-year course" – Kyle Duquaine

    - "Exploring the affective domain in introductory chemistry courses – A cluster analyses study" – Andrew Schuster

    - "Cluster analyses of the Chemistry Self-Efficacy Survey for CHEM 105 (Chemistry for Engineers)" – Evan Pagano

    - "Grading and analysis of Immediate Feedback Assessment Technique (IF-AT) responses" – Nicholas Vorwald

TEACHING EXPERIENCE

Teaching Assistant – Department of Chemistry and Biochemistry
University of Wisconsin – Milwaukee, WI

- Chemistry for Engineers – "***Flipped classroom***" model — Jan 2015 – May 2016
    - **Sole** TA for five discussions (~150 students)
- Chemistry for Engineers – "***Traditional classroom***" model — Sep 2013 – Dec 2014
- General Chemistry II — Sep 2012 – May 2013
- Preparatory Chemistry — Jan – May 2012

Teaching Assistant – Department of Chemistry
Purdue University – West Lafayette, IN

- General Chemistry I — Aug – Dec 2002
- General Chemistry II — Jan – May 2003
- General Chemistry for Engineers I — Aug – Dec 2003
- General Chemistry for Engineers II — Jan – May 2004
- Advanced General Chemistry (Honors course) — Aug – Dec 2004

Assistant Course Supervisor – Department of Chemistry Jan – Dec 2005
Purdue University – West Lafayette, IN

- Co-supervised 8 teaching assistants for General Chemistry II
- Conducted micro-teaching training for new teaching assistants
- Assisted course supervisor in developing handbook for teaching assistants
- Evaluated teaching assistants and provided feedback and support as required

OUTREACH

- Wisconsin Science Olympiad hosted by University of Wisconsin – Milwaukee – Volunteer
- GEMS (Girls Engineering Math and Science) – Volunteer
    - Presented poster on retention in STEM
    - Organized activities for kids to participate in e.g. online games, career quizzes
- MAGIC (More Active Girls In Computing) Mentor
    - Developed a chemistry project for a middle school student and mentored the student by providing guidance on experiments, writing a report and making a final presentation
- Volunteer – Junior Achievement, Moline High School, Moline, IL
- Volunteer – Literacy is for Everyone (LIFE), Blackhawk Community College, Moline, IL

National Chemistry Week – Purdue University, West Lafayette, IN