ASSESSING STUDENT SELF-REGULATION WITH A MODIFIED MICROANALYTIC APPROACH: INITIAL VALIDITY AND RELATIONS WITH STEREOETYPE THREAT

by

Faye C. Huie
A Dissertation
Submitted to the
Graduate Faculty
of
George Mason University
in Partial Fulfillment of
The Requirements for the Degree
of
Doctor of Philosophy
Education

Committee:

_________________________________________ Chair

__________________________________________

__________________________________________

__________________________________________ Program Director

__________________________________________ Dean, College of Education and Human Development

Date: ____________________________ Summer Semester 2015
George Mason University
Fairfax, VA
Assessing Student Self-Regulation with a Modified Microanalytic Approach: Initial Validity and Relations with Stereotype Threat

A Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at George Mason University

by

Faye C. Huie
Master of Education
George Mason University, 2008

Director: Anastasia Kitsantas, Professor
College of Education and Human Development

Summer Semester 2015
George Mason University
Fairfax, VA
This work is licensed under a creative commons attribution-noderivs 3.0 unported license.
Dedication

First and foremost I dedicate this work to my loving parents John and Mei Li, who have been a constant source of encouragement and affection since I began this journey. They are my life coaches and best friends. Every day, I am reminded of how lucky I am to have as wonderful parents as you.

To my little brothers Spencer, Chris, and Eric, who were the greatest burst of sunshine every time I felt this task was too daunting to accomplish. It is my greatest hope that my experiences inspire you all to push yourselves to your fullest potential, that no task is ever too challenging to undertake, and that no barrier is ever too large to overcome.

Above all, I dedicate this to my husband Savy Khuon. Without his love, cheesy sense of humor, and late night therapeutic conversations, I would not be where I am today. You are my rock and the source of my sanity. There is no question that the successful completion of this project is entirely due to having you by my side.
Acknowledgements

I wish to express my deep and sincere gratitude to all my professors who have helped shape my way of thinking and world view. This program has truly changed my life by teaching me to appreciate the intricacies of our social world. To my committee member Dr. Eden King, for providing me guidance on a topic that I never would have been able to sort out on my own. To Dr. Michelle Buehl, who has been undoubtedly the most exemplary professor and mentor. She is the person I try to emulate in both my teaching and research. Finally, my dissertation chair, Dr. Anastasia Kitsantas, who has been the driving force behind my work. She is both my good friend and greatest mentor. Thank you for believing in me and pushing me to complete this project. This would not have been possible without you.
**Table of Contents**

List of Tables .......................................................................................................................... viii
List of Figures .......................................................................................................................... ix
Abstract .................................................................................................................................. x

Chapter One ............................................................................................................................. 1
  Background of the Problem ................................................................................................. 1
  Statement of the Problem ................................................................................................... 2
  Purpose of the Study .......................................................................................................... 8
  Significance of the Study .................................................................................................. 9
  Definitions ........................................................................................................................... 10

Chapter Two ............................................................................................................................. 14
  Social Cognitive Theory .................................................................................................... 14
  Zimmerman’s (2000) Three Phased Cyclical Model of Self-Regulated Learning ....... 15
  Stereotype Threat ............................................................................................................... 24
    Distinguishing stereotype threat from related constructs ............................................. 26
    Causal relationship between stereotype threat and women’s achievement in math .. 28
  Self-Regulation Assessment Methodologies ................................................................. 33
  Microanalysis History and Background ....................................................................... 42
  Critical Components of SRL Microanalysis .................................................................. 45
  Validity of SRL Microanalysis .......................................................................................... 48
  Self-Regulation Assessment Tool: A Modified Microanalytical Approach .......... 50

The Relationship between Self-Regulation and Stereotype Threat .................................... 53
  Forethought phase: Self-efficacy ..................................................................................... 53
  Forethought phase: Attainment task value ..................................................................... 54
  Performance phase: Attention focusing strategies ......................................................... 56
  Performance phase: Metacognitive monitoring ............................................................... 59
  Self-reflection phase: Causal attributions ...................................................................... 62
Suggestions for future research ................................................................. 144
Appendix A .................................................................................................. 148
Appendix B .................................................................................................. 149
Appendix C .................................................................................................. 154
Appendix D .................................................................................................. 159
Appendix E .................................................................................................. 160
Appendix F .................................................................................................. 161
References .................................................................................................. 162
List of Tables

Table                                           Page

Table 1. SRAT:MMA Item Development Guidelines                      70
Table 2. Overall Self-Regulation and Self-Regulation by Achievement Level 106
Means and Standard Deviations                               
Table 3. Intercorrelations between all Individual Self-Regulation Variables 107
Table 4. Individual T-Tests with High/Low Self-Regulation as the Grouping Variable and Achievement as the Independent Variable 111
Table 5. Hierarchical Regression of all Individual Self-Regulation Variables entered into three Models Predicting Student Math Achievement 118
Table 6. Methodological Similarities and Differences Between Good Et Al. (2008) and this Present Study 121
List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1.</td>
<td>Zimmerman (2000) Three phased model of self-regulated learning</td>
<td>16</td>
</tr>
<tr>
<td>Figure 2.</td>
<td>Graph of the interaction between forethought and self-reflection phase composite variables by math test achievement levels</td>
<td>114</td>
</tr>
<tr>
<td>Figure 3.</td>
<td>Graph of the interaction between performance and self-reflection phase composite variables by math test achievement levels</td>
<td>115</td>
</tr>
</tbody>
</table>
Abstract

ASSESSING STUDENT SELF-REGULATION WITH A MODIFIED MICROANALYTIC APPROACH: INITIAL VALIDITY AND RELATIONS WITH STEREOTYPE THREAT

Faye C. Huie, PhD
George Mason University, 2015
Dissertation Director: Dr. Anastasia Kitsantas

The primary purpose of this research was to investigate the initial validity of a new approach to assessing student self-regulation based on microanalysis called the Self-Regulation Assessment Tool: A Modified Microanalytical Approach (SRAT:MMA). The secondary purpose was to explore whether self-regulation, as measured by the SRAT:MMA, may alleviate stereotype threat effects on women’s math achievement. A total of 191 undergraduate students enrolled in non-introductory math courses participated in the study. The classic microanalytic method was modified to be administered as a contextualized self-report survey to assess student self-regulation behaviors in real-time. The results revealed that students with high overall self-regulation tended to achieve higher than students with low self-regulation. In terms of the individual processes, students who reported higher levels of self-efficacy, math domain task value, metacognitive monitoring, stable attributions, and self-satisfaction achieved higher than
students who reported lower levels of those processes. Moderation analyses revealed that the relationship between the forethought and self-reflection phases and the relationship between the performance and self-reflection phases varied across levels of student achievement. The self-regulation processes altogether accounted for 70% of the variance in student achievement. Generally, the results may suggest that the SRAT:MMA is a valid alternative to current self-regulation assessment methodologies. In terms of stereotype threat, fidelity analyses showed that the manipulation failed to invoke stereotype threat in this specific sample. Overall, this study showed promising results regarding the initial validity of the SRAT:MMA. Implications of the findings and suggestions for further validation and research are discussed.
Chapter One

Self-regulation is a powerful determinant in student academic achievement, which emphasizes the self-motivating and sustaining role of the student in his or her own learning process (Zimmerman, 2000). Current methods of self-regulation assessment that take into account the dynamic nature of self-regulation lack the logistical and practical efficiencies required for large scale research and classroom application. Therefore, the primary purpose of the study was to investigate the initial validity of the Self-Regulation Assessment Tool: A Modified Microanalytic Approach (SRAT:MMA), which attempted to bridge the gap between assessment efficiency and theoretical soundness. The secondary purpose of this study was to investigate whether this new microanalytically modified method of measurement can be used to assess whether self-regulation can alleviate women’s experience of stereotype threat in math. Stereotype threat refers to the anxiety that one may experience when faced with a negative stereotype about his or her social group (Steele & Aronson, 1995). For example, women who are primed with the “men are better than women in math” stereotype will perform lower than men on math tests than women who were not primed with the stereotype (Schmader, Johns, & Barquissau, 2004).
Background of the Problem

Self-Regulation. Research on self-regulation emerged during the 1960’s and 70’s, and became popular in the 1980’s when theorists began to shift explanations for achievement from innate aptitude and advantageous environments to the role of self-controlled, self-directed processes (Winne & Perry, 2000; Zimmerman, 1989, 2000; Zimmerman & Schunk, 2011). According to this perspective, a highly intelligent student under the most ideal circumstances may still fail to achieve at high levels if the student does not possess the skills to control his/her environmental, affective, and motivational processes.

Since its conception, self-regulation and its individual processes has been a defining research topic in educational psychology. In fact, Boekaerts (1999) suggests that being able to self-regulate does not only help students achieve to their highest potential, but also in other aspects of life in general. In terms of academic achievement, however, these processes play a particularly important role (Zimmerman, 2008a). For example Pintrich and De Groot (1990) found that self-regulation positively predicted several different performance measures including: seat work performance, exams/quizzes, essays/reports, and average overall grades. Generally, findings regarding the positive impact of self-regulation on student academic achievement have been replicated in many different studies (Chen, 2003; DiBenedetto & Zimmerman, 2010, 2013; Kitsantas, 2002; Lynch, 2008; Schunk, Pintrich, & Meece, 2008; Zimmerman, Bandura, & Martinez-Pons, 1992), confirming that self-regulation is a highly important aspect in learning and achievement.
One of the main methods of assessing self-regulation since its inception was through retrospective self-report measures such as the Learning and Study Strategies Inventory (LASSI: Weinstein Schulte, & Palmer, 1990) and Motivated Strategies for Learning Inventory (MSLQ: Pintrich, Smith, Garcia, & McKeachie, 1993). Even though these instruments has been proven reliable and valid across studies (Duncan & McKeachie, 2005; Flowers, Bridges, & Moore, 2011; Zimmerman & Kitsantas, 2014), criticisms of self-report methods emerged when theorists proposed that self-regulation, essentially, is both an aptitude and an event (Cleary, 2011; Winne, & Perry, 2000; Zimmerman, 2008a), which cannot be accurately captured with self-report measures. Specifically, self-regulated learning is a dynamic process that varies with a specific context and task (Bandura, 1997).

The traditional self-report survey measures of self-regulation inherently assume that self-regulation is a fixed skill as opposed to a fluid process. For example, a student’s level of self-regulated learning likely depends on the larger context (e.g., self-regulated learning patterns may be different in math versus English) as well as the kind of tasks assigned within the context (e.g., written research paper versus a multiple choice exam). Overall, student self-regulatory profiles differ based on the context and tasks within the context. Therefore, measuring self-regulation as a global skill may capture an inaccurate picture of student self-regulation. In response to the limitations of self-report surveys, an alternative method of measuring self-regulated learning was developed called microanalysis (Bandura, 1982; Cleary, 2011; Zimmerman & Kitsantas, 1996, 1997).
Microanalysis involves measuring behaviors and cognitions as they occur within a given context and task (Cleary, 2011). The self-regulation microanalysis process usually involves collecting data one-on-one with a trained researcher questioning the student during key moments of the self-regulation process. This can be rather tedious and inefficient, especially in the context of large college classrooms and large scale research studies. Therefore, to encourage classroom professors as well as self-regulation researchers to accurately assess students’ self-regulated learning, this study attempted to develop a modified student self-report survey based on microanalytic principles.

**Stereotypes.** Stereotypes have been a topic of social psychology research as early as 1933, when Katz and Braly published an article in the Journal of Abnormal and Social Psychology asking one hundred Princeton college students to indicate the characteristics of different social groups. The results showed that students generally agreed that Jews were “shrewd” (79%), Japanese were “intelligent” (47.9%), and African Americans were “lazy” (75%). Following their study, Gilbert (1951) and Karlins, Coffman, and Walters (1969) have replicated their findings with similar results known all together as the Princeton Trilogy (Devine & Elliot, 1995). Findings from the Princeton Trilogy propelled researchers into examining various forms of social stereotyping including implicit stereotyping (Gross & Hardin, 2007), stereotype vulnerability (Aronson & Steele, 2005), stigma consciousness (Pinel, 1999), and stereotype threat (Steele & Aronson, 1995).

Most of the research within the stereotype literature has been conducted within the social psychology field. However, there are clear implications of how stereotyping can influence student learning achievement. For example, in 1995, Steele and Aronson
published what is now considered a classic article that introduced the concept of stereotype threat. In their study, they proved that stereotypes play a powerful role in determining African American student academic achievement and intelligence outcomes. Specifically, Steele and Aronson describe stereotype threat as a predicament, where negative stereotypes about one’s own social group can cause: a) others to expect those individuals to confirm those stereotypes and b) inadvertently cause the individual to confirm the stereotype themselves. This predicament is true for race (Steele & Aronson, 1995) and gender (Rydell, Rydell, & Boucher, 2010) particularly in the area of educational achievement where stereotypes regarding intelligence within a particular domain are salient (e.g., African Americans are not as intelligent as Caucasians and men are better than women in math).

Steele and Aronson (1995) argue that negatively stigmatized individuals are at risk for confirming their social groups’ negative stereotype and experience an elevated level of fear and anxiety of conforming to the stereotype. Evidence to support this argument was obtained through two experimental studies investigating the stereotype that African Americans are generally not as “smart” as their Caucasian counterparts. The findings revealed that, first, African American students performed significantly lower than Caucasian students on a standardized test when negative stereotypes regarding African American intelligence was activated. However, achievement differences between African American and Caucasian students were not detected when that same test was described as not diagnostic of intelligence. Second, African American students in the high threat condition experienced elevated levels of concern regarding their ability,
disassociated themselves with African American stereotypes, and made more statements that would excuse their performance on the test than students in the low threat condition. Thus, with this study, stereotype threat is defined as “that in situations where the stereotype is applicable, one is at risk of confirming it as a self-characterization, both to one's self and to others who know the stereotype” (Steele & Aronson, 1995, p. 808). In other words, stereotype threat is the fear of confirming a negative stereotype about one’s own social group.

Statement of the Problem

During the 1990’s and early 2000’s, researchers conducted a series of studies investigating the relationship between self-regulation and learning (Cleary & Zimmerman, 2001; Kitsantas & Zimmerman, 2002; Kitsantas, Zimmerman, & Cleary, 2000; Zimmerman & Kitsantas 1996, 1997). In their studies, student self-regulation was measured in an authentic setting, where they were questioned at key points of the self-regulated learning process. Results of those studies revealed that this method of data collection is not only accurate and reliable, but it can detect performance level differences based on self-regulation levels. Although the methods used in some of those studies were not explicitly described as “microanalytic,” it set the framework for developing a more comprehensive method of microanalysis used today.

Although the theoretical and empirical advantages of microanalysis are clear (DiBenedetto & Zimmerman, 2013), this method has yet to gain popularity, which may be due to the fact that it is time consuming and requires a large amount of resources. Cleary (2011) suggests that a comprehensive microanalysis is characterized by five major
components: 1) individualized assessment, 2) assessment of multiple self-regulation processes, 3) context-specific questions, 4) alignment between questions and the three phased cycle of self-regulation, and 5) response recording and coding. Based on Cleary’s (2011) guidelines, the most time-consuming nature of this form of assessment is the one-on-one interaction required between the researcher and student. That is, in order for a method to be described as truly microanalytic, the researcher must be present with the student throughout an entire self-regulation cycle, making the microanalytic process an impractical method for larger scale research studies or even teachers who are interested in diagnosing their students’ self-regulation. Thus, the problem with regards to the microanalytic method is how it can be modified to fit the real-world needs of researchers and teachers while still maintaining the integrity of the method?

Stereotype threat is a highly contextualized phenomenon that is difficult to empirically detect in naturalistic settings (Nguyen & Ryan, 2008). However, its impact on educational achievement has been replicated in hundreds of laboratory studies, confirming that this phenomenon not only exists, but it is also contributing to the gender and racial achievement gap (Aronson & Inzlicht, 2004; Aronson & Steele, 2005; Good, Aronson, & Harder, 2008; Spencer, Steele, & Quinn, 1999; Steele & Aronson, 1995). These gaps continue to persist despite interventions and programs designed to curb these statistics (Good, Aronson, & Inzlicht, 2003). If stereotype threat is highly contextualized, then it would seem sensible that the mediators of the threat are also highly contextualized. Therefore, the microanalytic method, which is inherently designed to measure processes as they occur in a given context, may be an appropriate method of
identifying the potential mediators of this effect. Therefore, if stereotype threat is a real problem that is contributing to the current achievement gap, and if social psychologists have struggled to identify the mediators of stereotype threat, then concepts in educational psychology has the potential to explain stereotype threat and offer methods for instructors to ameliorate this predicament.

Thus, the overall problem was twofold: 1) how can educational psychologists modify the microanalytic method to accurately and efficiently measure student self-regulation; and 2) how this method based in microanalysis can be implemented to identify the self-regulation factors that may lessen the impact of stereotype threat among traditionally stigmatized groups, namely, women in math.

**Purpose of the Study**

The major purpose of this study was to explore the initial validity of the SRAT:MMA. The secondary purpose of this study was to investigate whether this method of assessment can be used to explore the effects of self-regulation on stereotype threat in undergraduate math students. Overall, microanalysis was designed to more precisely assess self-regulation as they occur under specific conditions. Stereotype threat, by nature, is difficult to capture and occurs under very specific conditions (Nguyen & Ryan, 2008). Therefore, this study attempted to not only develop a method of assessing self-regulation that combines the efficiency of a student self-report measure with the theoretical soundness and accuracy of microanalysis, it also attempted to apply this measure in a meaningful context (i.e., taking a math test) under specific conditions (i.e., high and low stereotype threat conditions).
Significance of the Study

The theoretical advantages of microanalysis are that: it takes into account that self-regulation is both an aptitude and event; self-regulation is different under different contexts; and that self-regulation is a dynamic process. Therefore, the results of microanalysis capture a more accurate picture of student self-regulation than other popular methods such as self-report surveys because its principles directly align with self-regulation theory. However, even with these advantages, performing a full microanalysis is logistically difficult and inefficient. Therefore, the practical significance of this study is that: if the results of this study show that self-regulation can be accurately assessed through a modified, more efficient form of microanalysis, then this study may yield a promising alternative to traditional self-report measures of self-regulation for both teachers and researchers.

In terms of stereotype threat, since the inception in 1995 by Steele and Aronson, stereotype threat has been one of the most popularly studied phenomena in social psychology (Schmader et al., 2004). However, the mechanisms that mediate the relationship between stereotype threat and achievement is not clear (Aronson & Steele, 2005; Ryan & Ryan, 2005) and no research to date have examined this effect holistically through the lens of social cognitive theory. This may be due to the fact that stereotype threat is heavily nuanced and dependent on the context (Nguyen & Ryan, 2008). Therefore, by using a context-specific measure of self-regulation within a stereotype threat environment, the theoretical significance of this study is that it attempts to merge together two fields: social psychology (e.g., stereotype threat) and educational
psychology (e.g., social cognitive theory and self-regulated learning) in an effort to explain stereotype threat effects.

The following research questions were investigated:

1. Can the SRAT:MMA diagnose college student self-regulated learning at the classroom level during test taking in a math course?
   a. Do students with high self-regulation perform better than students with low self-regulation on a math test?
   b. Does the relationship between the different phases of self-regulation vary across levels of student achievement on the math test?
   c. Can the self-regulation variables, as measured by the SRAT:MMA predict college student achievement on a math test?

2. Can self-regulation processes, as measured by the SRAT:MMA, alleviate stereotype threat effects in women enrolled in undergraduate math courses?

Definitions

**Stereotype threat**: Stereotype threat refers to the anxiety that one experiences when faced with a negative stereotype regarding their social group (Steele & Aronson, 1995).

**Social cognitive theory**: Social cognitive theory asserts that individual behaviors and cognitions are the result of three interrelated factors: personal, behavioral, and environment (Bandura, 1989).

**Self-regulation**: Self-regulation refers to the degree to which students can proactively and metacognitively engage in different behavioral, motivational, and
cognitive processes to attain a self-set goal. Zimmerman (1989) suggests that students can self-regulate to the extent that they can cyclically engage in three different phases of learning: forethought, performance, and self-reflection.

**Forethought Phase:** This is the first phase of the three phase cyclic model of self-regulation. Students engage in the processes under this phase prior to the learning task which help prepare them to learn. Processes under the forethought phase fall under two categories: task analysis and self-motivational beliefs.

A. **Task analysis:** The analysis and breakdown of a specific learning task to determine the most effective completion method.

   a. **Goal setting:** Goal setting is the process through which students determine the type of outcome that they wish to achieve (Locke & Latham, 1990). Characteristics of goals can include a high degree of difficulty, and process oriented goals versus outcome oriented goals (focus on the processes of achievement as opposed to the actual outcome itself, which are outcome oriented goals) (Zimmerman, 2008b).

   b. **Strategic planning:** Strategic planning refers to how students decide on which strategies to use while performing the task to successfully achieve goals (Zimmerman, 2000).

B. **Self-motivational beliefs:** This refers to the system of belief formed by students to successfully initiate, sustain, and complete the learning task. The processes under this system include:
c. Self-efficacy: The degree to which students feel capable of successfully accomplishing a certain task (Bandura, 1989).

d. Attainment task value: Perceptions regarding how personally important it is to excel at a certain task or domain (Eccles & Wigfield, 1995).

**Performance Phase:** This is the second phase of Zimmerman’s (1989) three phase cyclic model of self-regulated learning. Processes under this phase occur during student learning efforts and fall under two main categories: self-control and self-observation.

A. Self-control: This refers to how well students can control their efforts to learn (Cleary & Zimmerman, 2004). A process under this category of self-regulation include:

   a. Attention focusing strategies: Strategies that improve the students’ ability to avoid distractions and to focus on the task at hand (Zimmerman, 2008a).

B. Self-observation: This refers to how students can track, monitor, and reflect on their progress while performing a task to achieve desired outcomes and goals (Zimmerman & Kitsantas, 2005).

   a. Metacognitive monitoring: The extent to which students are aware of their comprehension and progress while performing a task (Nietfeld, Cao & Osborne, 2005; Schunk, 1996).

**Self-Reflection:** This is the third phase of Zimmerman’s (1989) three phase model of self-regulated learning. Processes under this phase occur after the task is complete and fall under two major categories: self-judgment and self-reaction. These processes
ultimately influence other factors under the forethought and performance phase, hence, the cyclic nature of self-regulated learning.

A. Self-judgment: This refers to how students self-evaluate their learning and assign causal attributions to their achievement outcomes.

   a. Self-evaluation: The judgement criteria used to assess learning and performance by specific mastery criteria, past performance, or social comparison (Cleary & Zimmerman, 2004).

   b. Causal attributions: The explanations to explain a particular outcome. Specifically, Weiner (1986) suggests that attributions can be categorized under three general causal dimensions: controllable (effort) or uncontrollable (luck), stable (ability) or unstable (effort), and internal locus of control external (ability) or external locus of control (luck).

B. Self-reaction: Self-reaction is the category of processes that describe the affective reactions to the outcomes of the learning task.

   a. Self-satisfaction: Positive or negative emotional reactions to the learning outcomes (Zimmerman, 2008a).
Chapter Two

The following sections will review the literature surrounding self-regulation, methods of assessing self-regulation, and evidence of the relationship between self-regulation and stereotype threat. The literature review will begin with a discussion of self-regulation and stereotype threat. The second section will delve deeper into the current methods of assessing self-regulation, focusing on microanalysis and how it was modified for the purposes of this study. Finally, the third section will describe the links between self-regulation and stereotype threat, and why a microanalytic methodology for examining the relations between these two concepts is necessary.

Social Cognitive Theory

Zimmerman’s (2000) theory of self-regulation is grounded in social cognitive theory (Bandura, 1986). Social cognitive theory was developed by Albert Bandura in the 1980’s as an explanation for human functioning. Specifically, this theory proposes that human learning involves the interaction between behavior, personal cognitive processes, and the environmental context, known as the triadic reciprocality. This model of learning suggests that all three processes interact with each other to ultimately determine outcomes. Specifically, student behaviors and achievement is dependent on the motivational and affective perceptions (personal cognitive variables) coupled with the
social context (environmental factors). According to this perspective, a student can achieve and perform to the extent to which he/she can successfully regulate these factors.

**Zimmerman’s (2000) Three Phased Cyclical Model of Self-Regulated Learning**

Social cognitive theory can be further conceptualized as self-regulated learning. Based under the larger umbrella of social cognitive theory, self-regulated learning is defined as the ability for a student to strategically, proactively, and independently engage in thoughts and behaviors to attain self-set, personal goals (Zimmerman, 1989). Self-regulation was further conceptualized when Zimmerman (1989) developed a three phased theory of learning to explain how students engage in different processes to independently sustain their efforts to learn. The main assumption of this theory is that all students are able to achieve to their highest potential to the degree that they are able to successfully regulate three phases of learning. These sequential and cyclically related phases of learning are: forethought, performance, and self-reflection/evaluation (See Figure 1). Students engage in the forethought phase prior to actually engaging in the task at hand, where they prepare to learn. The performance phase occurs when the student is engaged in actual learning, or completing an academic task. The third and final phase is self-reflection, where students react and respond to their academic outcomes. Critical to this perspective is the cyclic feedback loop. Specifically, this feedback loop provides the student with information about how to best adapt ones behaviors and thoughts to more successfully attain personal goals.
Forethought phase. The forethought phase of self-regulation involves affective and motivational processes which fall under two categories: task analysis and self-motivational beliefs (Zimmerman, 2000). Task analysis refers to how students analyze the learning task prior to performance such as goal setting and strategic planning. Self-motivational beliefs refer to the beliefs that students' form to successfully initiate and sustain learning efforts such as self-efficacy and task value.
Goal setting and strategic planning are two integral components of the forethought phase. Goal setting not only directs a student’s behavior and plan of action towards attaining the goal, but it also influences their level of motivation (Zimmerman, 2008b). Specifically, students who set challenging, specific, proximal, and process oriented goals tend to exert more effort in attaining those goals as well as persist in the face of difficulty than those who do not (Zimmerman, 2008b). Due to these motivational influences of goal setting, engaging in this behavior positively influences students’ level of academic achievement (Bandura & Schunk, 1981; Locke & Latham, 1990; Zimmerman, 2000; Zimmerman & Kitsantas, 1997, 1999; Zimmerman & Martinez-Ponz, 1986, 1988). Strategic planning refers to how students develop a “plan of action” to achieve their goals (Zimmerman, 2008a). Students who actively and effectively set goals plan out the strategies for guiding cognition, controlling affect, and directing learning-related behaviors (Winne, 1997). However, both goal setting and strategic planning are influenced by their motivational beliefs in their ability to effectively execute and attain their goals.

Self-efficacy and task value are two critical factors underlying the self-motivational beliefs category of the forethought phase. These factors guide the degree to which students engage in effective goal setting and strategic planning. Specifically, self-efficacy refers to one’s belief in their ability to successfully accomplish a certain goal (Bandura, 1997). Decades of educational research has established self-efficacy as a powerful factor in determining students’ level of academic achievement (Schunk, 1991). Specifically, there is evidence that self-efficacy is a key motivational belief that
influences students’ effort and persistence in completing tasks (Shell & Husman, 2008) and the use of effective self-regulatory learning strategies (Bandura, 1997). In a meta-analysis of 109 studies, Robbins et al. (2004) found that out of nine constructs, academic self-efficacy was the strongest predictor of GPA. Similarly, in another meta-analysis, Multon, Brown, and Lent (1991) found that academic accomplishments significantly and positively correlated with positive self-efficacy beliefs. Task value, on the other hand, is another factor under the self-motivational beliefs category of the forethought phase.

According to Eccles and Wigfield, (1995) task value is composed of four constructs: attainment, intrinsic, utility, and cost. Attainment task value refers to how personally important it is for the individual to do well in a given task or domain whereas intrinsic value is the degree of enjoyment one gains from engaging in the task. Utility value is how useful the task is in terms of fulfilling future goals (e.g., the ideas taught in class are relevant to the students’ future career goals) and finally cost, which refers to the sacrifices one must give to complete the task. In terms of stereotype threat, the most relevant aspect of task value is attainment task value (Aronson, Lustina, Good, Keough, & Steel, 1999; Keller, 2007). Eccles (2005) suggests that attainment task value is closely related to identity. Students with high attainment task value generally view the skills associated with completing the task as an important component of their identity. For example, a student who has high attainment task value in math would view his or herself as a “math person.” Doing well in math would then confirm important aspects of themselves—or their identity. Generally, in educational psychology, research shows that having a high attainment task value is related to higher levels of academic achievement.
A similar concept in social psychology is domain identification, which refers to the degree to which an individual identifies with a certain domain (e.g., math is an important part of an individual’s identity). However, in the stereotype threat literature, the more domain-identified an individual is, the more susceptible he/she is to stereotype threat, whereas in educational psychology, the more domain-identified an individual is, the higher he/she will achieve academically.

**Performance phase.** After students have formed their motivational and affective beliefs in the forethought phase, learners then proceed to the performance phase where they actually engage in accomplishing the task/goal at hand. Processes under this phase are categorized into self-control and self-observation strategies. Self-control is the behavioral and cognitive processes by which students engage in to enhance their learning and achievement such as attention focusing strategies. Self-observation is the second category of processes that refers to the strategies that students engage in to observe and monitor their learning behaviors, such as metacognitive monitoring (Zimmerman, 2008a).

An important self-control process is attention focusing strategies (Cleary & Zimmerman, 2006; Zimmerman, 2000). Attention focusing strategies are the behaviors that students engage in to sustain their efforts and motivation to complete a learning task (Cleary & Zimmerman, 2006). Specifically, strategies such as structuring and managing the learning environment to minimize distractions (e.g., time management, effort regulation, planning) help the student in attaining academic goals while maladaptive strategies may include avoidance of the task or poor management abilities.
Metacognitive monitoring, or the thoughtful, active monitoring of one’s own thought process (Metcalfe, 1986; Metcalfe, Schwartz, & Joaquim, 1993; Schraw, 1994), is important not only for student academic achievement, but also for developing self-regulated learning (Butler & Winne, 1995). Metacognitive monitoring is an executive process (Moores, Chang, & Smith, 2006; Nietfeld et al., 2005) that includes specific judgements of progress and calibration of performance. For example, a metacognitive judgement could be, “I’m pretty confident I got at least the last three questions right—at the rate I’m going I’m sure that that I am going to get at least a B on this test.” This type of metacognitive monitoring shows that students are aware of what they know and do not know. Students are particularly skilled at metacognitive monitoring if their judgements are accurate. In the previous example, if the student actually did receive a B on the test, it would indicate strong metacognitive monitoring skills.

Overall, students who are skilled at metacognitive monitoring and who are more accurate in their metacognitive judgments tend to achieve higher than students who are less skilled at metacognitive monitoring (Nietfeld et al., 2005). In fact, Nietfeld et al. (2005) showed that students who had more specific (performance on a specific test) rather than general (overall GPA) knowledge were more able to successfully engage in metacognitive monitoring, which ultimately led to higher academic achievement than students who were less able to make accurate metacognitive judgments. In other words, students who have stronger knowledge in specific academic areas (knowledge of the content in a math test as opposed to more general knowledge of math) tend to show
stronger levels of metacognitive monitoring, which result in higher academic achievement.

**Self-reflection phase.** The final phase of self-regulated learning is self-reflection, which includes processes that allow the student to self-judge his/her performance. Based on the self-reflective information, the student then makes changes to their behaviors/cognitions in the forethought and performance phase in order to more successfully complete future tasks (i.e., cyclical feedback loop). The processes under this phase are separated into two major categories: self-judgment and self-reaction. Self-judgment refers to the processes that students engage in to self-evaluate their learning outcomes as well as make causal attributions to their successful and unsuccessful outcomes. Self-reaction refers to the affective reactions that students have regarding their achievement outcome, such as self-satisfaction (Zimmerman & Kitsantas, 1999).

After students have received the outcome (e.g., performance on a test), students make self-evaluative judgments about their performance. Specifically, students compare their performance to a self- or other-imposed standard or goal. Proactive, highly self-regulated students use standards set in the forethought phase to evaluate their outcomes whereas reactive, low self-regulated students either fail to self-evaluate or compare themselves against their peers (Zimmerman, 2008a). Research shows that self-evaluation is a critical process linked to the cyclic nature of self-regulated learning. That is, proactive learners self-evaluate based on personal goals set in the forethought phase and then proceed to make adjustments to behaviors in the performance phase to make learning more effective. For example, Schunk and Ertmer (1999) investigated how goal
setting and self-evaluation influenced college students’ level of self-efficacy, achievement, and self-efficacy to engage in self-regulatory behavior. Students who set process goals and evaluated their own learning throughout the studying process were more likely to have higher self-efficacy, strategy use, and were more skilled than students who did not engage in self-evaluation.

Another important component of the self-reflection phase is called attributions, which refers to the perceived causes of a particular outcome (Weiner, 1986). Weiner (1986) suggests that attributions typically fall along three causal dimensions which are: stability (dynamic or static over time and events), controllability (controllable or uncontrollable), and locus of control (internal or external). For example, an attribution such as effort would be an unstable/internal/controllable attribution whereas an attribution such as luck would be an unstable/external/uncontrollable attribution. Although different attributions to success or failure have different behavioral and psychological consequences, generally, adaptive attributions are those that are unstable, internal, and controllable while maladaptive attributions are those that are stable, external, and uncontrollable (Schunk et al., 2008). Generally, students tend to achieve higher and take more control over their learning when they perceive that they are responsible for their own learning and academic outcomes. For example, a student who attributes their poor performance on a math test to a lack of effort is more likely to engage in self-regulated learning than a student who attributes their poor performance to innate low mathematical ability. However, there are circumstances where students who make more maladaptive attributions continue to achieve high academically.
Although research generally shows that controllable attributions are most predictive of student academic success, Gibb, Zhu, Alloy, and Abramson (2002) found that college students with high SAT scores who made maladaptive attributions (internal or stable attributions) for negative outcomes received higher cumulative college GPAs than students with low SAT scores and who made maladaptive attributions. This may indicate that maladaptive attributional styles negatively influence student achievement only if the student has not traditionally achieved high academically. Furthermore, Gibb et al. (2002) found that freshmen students with adaptive attributional styles (external or unstable attributions for negative events) earned approximately the same cumulative GPAs regardless of high or low SAT scores. This suggest that although maladaptive attributions negatively influence students who have traditionally not achieved high, students who have more adaptive attributional styles do well academically regardless of past achievement.

Finally, self-satisfaction is a subprocess under self-reactions, which involves determining how satisfied one is with a particular outcome. Overall, self-satisfaction influences other aspects of self-regulation like motivation and effort regulation (Zimmerman & Kitsantas, 1999) in addition to influenced by aspects such as attributions (Weiner, 1986). For example, a student who attributes his or her poor performance to “I’m just no good at math, it’s just too hard for me” is more likely to experience stronger dissatisfaction, which would in turn negatively influence his or her motivation and effort in the forethought and performance phases of learning. As a result, students’ satisfaction
is an important self-regulation process because it plays a vital role in determining the interrelationships between phases (DiBenedetto & Zimmerman, 2013).

Overall, the literature concludes that self-regulation plays an important role in a students’ level of academic achievement. The three-phased self-regulation model is particularly important, given that learners sustain their own learning by setting personal goals and engaging in a self-oriented feedback loop. Engagement in the feedback loop allows students to monitor their own progress and adapt their behaviors accordingly. It is also important to note that although self-regulation is an independent form of learning and that these processes are largely influenced by the social environment (Zimmerman, 2008a). Stereotypes, which are a product of the social environment, may influence the degree of self-regulated learning one engages in or, by contrast, self-regulated learning may influence the degree to which individuals cope with stereotypes. These next sections will provide a brief overview of stereotype threat, differences between stereotype threat and other related constructs such as stereotype priming and stigma consciousness, as well as the empirical evidence regarding the causal relationship between stereotype threat and achievement.

**Stereotype Threat**

Stereotype threat was coined by Steele and Aronson in 1995 in an attempt to explain the relationship between situational stereotypes and individuals’ performance or behaviors. Specifically, Steele and Aronson (1995) suggest that individuals experience a threat when faced with a negative stereotype against their social group. In order for the threat to be experienced, however, one must only be aware of the stereotype. In other
words, the individual must be aware that within that given context, there is a stereotype against his/her social group. Take for example a female student completing a math intelligence test to help researchers understand why males tend to achieve higher than females in math. In order for stereotype threat to occur, the female student must only know that the stereotype: “men are better than women in math” exist. Even if the student does not believe that this stereotype is true, it will not protect her against the effects of stereotype threat (Steele & Aronson, 1995). Stereotype threat is detected and confirmed when individuals who are targets of the stereotype underperform as compared to others who are not targets of the stereotype.

It is also important to note that the stereotype threat effect is not limited to a specific type of stereotype. That is, individuals who are the target of any negative stereotype are vulnerable to its effects (Steele & Aronson, 1995). Research has shown that individuals who are subjected against gender (Schmader, Johns, & Forbes, 2008), socioeconomic status (Croizet & Claire, 1998), athletic ability (Stone, Lynch, Sjomeling, & Darley, 1999), and ethnicity/race, including Caucasian men (Aronson et al., 1999) all carry the burden of stereotype threat. The focus of this study was on stereotype threat with regards to gender and math.

Research has confirmed that in order for stereotype threat to occur, the stereotype must be relevant in the specific context that the individual is in. Specifically, stereotype threat is a situational phenomenon (Good et al., 2008; Schmader et al., 2008). However, it is important to distinguish stereotype threat from other similar constructs.
**Distinguishing stereotype threat from related constructs.** Although stereotype threat is a popular and unique concept that measures the repercussions of social stereotypes, several other similar constructs exists. These include, but are not limited to: stereotype priming, stigma consciousness, and stigma schematicity. The following sections will compare and contrast stereotype threat with those other concepts.

Stereotype priming is defined as when a specific stimulus implicitly activates prior knowledge associated with that stimulus, in turn, influencing the performance and behaviors of the individual (Marx, 2011). For example, in a classic stereotype priming study, Bargh, Chen, and Burrows (1996) found that college students who were primed with elderly stereotypes reported that the elderly tended to walk slower than students who were not primed with elderly stereotypes. Stereotype priming and stereotype threat both have the same effect on performance (Marx & Stapel, 2006), however, they are not interchangeable. In order for stereotype threat to occur, the individual must both be cognizant of the stereotype as well as be a member of the social group that the stereotype is targeted towards. This is in comparison to stereotype priming, where an individual must only be aware of the stereotype for priming to occur, regardless of whether or not they are part of the stereotyped social group. Stereotype priming is therefore a more general construct where it can affect anyone as opposed to stereotype threat, which is a more specific construct that can only affect those who are both knowledgeable and targets of the stereotype.

Stigma consciousness refers to the extent to which an individual expects to be stereotyped by others, regardless of the actual behavioral outcome (Pinel, 1999).
Although stigma consciousness and stereotype threat do covary, Steele (1997) suggests that stereotype threat affects behaviors and/or performance only when individuals are placed in a situation where those stereotypes are activated as opposed to sigma consciousness, where the expectations of being stereotyped is the main concern. Those who experience high levels of stigma consciousness tend to express more concern about being stereotyped than those who experience lower levels of stigma consciousness (Brown & Pinel, 2003). With this line of reasoning, we can infer that those who are high in stigma consciousness would be especially vulnerable to stereotype threat effects.

Stigma schematicity (Jones et al., 1984) is another similar construct that is defined as the degree of internalization that individuals have regarding the different stereotypes of their social group. For example, an Asian male who is stigma schematic would believe that he is smart in math because he is Asian and is a male. In comparison to stereotype threat, individuals experiencing stereotype threat may or may not internalize the stereotype regarding their social group. Specifically, individuals need not be stigma schematic in order to experience the effects of stereotype threat.

Although all of these similar constructs may act as moderators and increase individuals’ vulnerability to stereotype threat, none are required in order to detect stereotype threat. Stereotype threat is confirmed when individuals exposed to a threat actually underperform as a result of that exposure. That is, a behavioral or performance outcome is measured against a group of individuals who were not exposed to the stereotype threat. However, other similar constructs such as stigma consciousness and group identification do not necessarily focus on a certain outcome, where the expectation
(i.e., stigma consciousness), and internalization (i.e., stigma schematicity) is the primary focus. Another critical difference between these constructs and stereotype threat is the degree of variability measured. Specifically, unlike stigma consciousness and stigma schematicity, individual perceptions in stereotype threat are not assessed—stereotype threat is detected only when a group exposed to the threat performs significantly lower than another group that was not the targets of the threat.

Now that the differences between stereotype threat and other similar constructs have been contrasted, the following section will review research that establishes the causal relationship between stereotype threat and achievement, specifically women in math.

**Causal relationship between stereotype threat and women’s achievement in math.** Prior research has established a causal effect between women achievement in math and stereotype threat (Brown & Josephs, 1999; Croizet et al., 2004; Dar-Nimrod & Heine, 2006; Davies, Spencer, Quinn, & Gerhardstein, 2002; Good et al., 2008; Inzlicht & Ben-Zeev, 2000; Johns, Schmader, & Martens, 2005; McGlone & Aronson, 2006; McIntyre, Paulson, & Lord, 2003; Schmader, 2002; Schmader & Johns, 2003; Schmader, Johns, Keifper, Healy, & Fairchild-Ollivierre, 2001; Spencer et al., 1999; Stoet & Geary, 2012). The first study to investigate the role of stereotype threat in women’s math performance was Spencer et al. (1999). Specifically, Spencer et al. (1999) designed three studies to investigate the role of stereotype threat in women’s math performance. The first study sought to establish the causal relationship between the difficulty level of the math test and gender differences. That is, stereotype threat is more pronounced on
difficult math tests as opposed to easy math tests. To confirm this hypothesis, Spencer et al. (1999) randomly administered either high or low difficulty math tests to 28 men and 28 women college students. The results showed that men scored higher than women on the difficult math test while women scored equally as men in the easy math test.

To further confirm that stereotype threat was the factor causing these differences (as opposed to men actually having stronger math skills than women), Spencer et al. (1999) conducted a second study. In the second study, all students were administered the difficult math test and conditions varied according to stereotype threat (high versus low stereotype threat). The assumption here was that if the results in Study 1 were due to actual ability differences between men and women on difficult math tests, then women should continue to underperform on difficult math tests regardless of how relevant a stereotype is. The results of the second study showed that women and men achieved at similar levels in the low stereotype threat condition whereas men achieved higher than women in the high stereotype threat condition. This further confirmed that stereotype threat was the factor that caused women’s math underperformance, not ability. In the third study, the authors attempted to reconfirm the stereotype threat effect on women’s math performance as well as investigate potential mediators (i.e., evaluation apprehension, state anxiety, and self-efficacy). Although the results again confirmed the stereotype threat effect, it also revealed that self-efficacy, evaluation apprehension, and anxiety did not successfully mediate the relationship between stereotype threat and achievement.
Spencer et al. (1999) very clearly established the causal relationship between stereotype threat and women’s math performance. This study has since propelled many researchers to both replicate this effect and identify the psychological and contextual constructs that may explain this relationship. For example, Inzlicht and Ben-Zeev (2000) investigated whether the context to which individuals are placed in can influence problem solving behavior. Specifically, Inzlicht and Ben-Zeev (2000) hypothesized that female college students experience of stereotype threat in math would be lessened in a context where females were the majority in the test taking room as opposed to the minority (i.e., three females in the test taking room versus one female and two males). In other words, women who take a difficult math test in the presence of men would experience stereotype threat effects whereas women who take a difficult math test in the presence of women would not experience stereotype threat effects.

To investigate this research question, Inzlicht and Ben-Zeev (2000) randomly assigned 72 females to a female only or female minority condition. Two tests were administered: a verbal and math test. The researcher informed the female participants that the purpose of the study was to investigate a training program for performance on standardized achievement tests and that their performance would be reported to other groups. The results showed that female students in the female only condition had achieved higher math scores than the female students in the female-minority condition. In fact, women’s math performance decreased as the number of men in the group increased. Further, this effect was only detected for the math test as opposed to the verbal test, where both males and females in both conditions scored equally. This study revealed that
stereotype threat exists more so in an environment that threatens an individual’s intellectual performance as opposed to an environment that is less threatening to an individual’s intellectual performance. Women tend to be more aware of gender stereotypes when outnumbered by men, which unfortunately is typical in upper level college math classrooms.

College women who are highly skilled in math and are in a pipeline to becoming professionals in the math and science fields are not spared the effects of stereotype threat. Specifically, Good et al. (2008) hypothesized that college women enrolled in rigorous mathematic programs would perform significantly lower than men in a stereotype threat condition than women who were not exposed to stereotype threat. A total of 157 (57 female; 100 male) students participated in the experiment. The results showed that females in the high stereotype threat condition had performed significantly lower than males whereas females in the low stereotype threat condition actually outperformed their male counterparts as well as answered more questions. However, even though female students in the low-threat condition had outperformed their male counterparts, they reported lower levels of confidence in their answers. The Good et al. (2008) study provided compelling evidence that: 1) even the brightest, most motivated women in math will experience stereotype threat effects; and 2) when the threat is lifted, females can outperform their male counterparts.

These studies suggest that activating a negative stereotype in a domain valued by the individual interferes with his or her ability to perform. Ultimately, there is evidence to strongly conclude that stereotype threat plays a significant and negative role in women’s
math achievement. In fact, since the seminal piece by Spencer et al. (1999), stereotype threat has been frequently cited as a potential explanation for gender differences in math performance (Stoet & Geary, 2012). Although researchers have attempted to investigate the moderators and mediators of the stereotype threat effect, the underlying cognitive and behavioral processes have yet to be thoroughly understood (Rydell et al., 2010; Schmader et al., 2008), particularly with women and math. Although Schmader (2002) has concluded that gender identification significantly moderates the relationship between women’s math performance and stereotype threat, researchers are still not clear on what can be done to combat stereotype threat effects.

From the social cognitive perspective, it is clear how this can be used as a framework to understand the effects of social phenomena on student achievement. Specifically, if stereotypes are social in nature and if social cognitive theory posits that achievement is the result of the social environment and individual personal factors, then social cognitive theory has the potential to explain the causal effects of stereotypes on achievement. From the research reviewed thus far, it is clear that both self-regulation and stereotypes are highly context specific (Cleary & Zimmerman, 2004; Cleary, 2011; Good et al, 2008; Inzlicht & Ben-Zeev, 2000; Nguyen & Ryan, 2008). As a result, measuring self-regulation within stereotype threat would require highly contextualized methods. Therefore, the purpose of the following sections will review the current methods of assessing self-regulation as well as the issues regarding the different methods.
Self-Regulation Assessment Methodologies

Self-regulation assessment has been an important topic since its conception in the 1980’s (Zimmerman & Martinez-Pons, 1986, 1988). Since then, several methods of measuring self-regulation have been developed. These include self-report surveys, interviews, direct observations, think-alouds, diaries, trace methods, and microanalysis. These following sections will discuss the strengths and limitations of each form of measurement as well as introduce a new form of assessing self-regulation.

Self-report surveys. Several survey instruments were developed in the 1980’s to assess students level of self-regulated learning. Since then, self-report surveys are currently the most commonly used methods of assessing self-regulation (Cleary, 2011; DiBenedetto & Zimmerman, 2013; Perry & Winne, 2006; Wolters, Benzon, & Arryo-Giner, 2011). The popularity of self-report surveys stems from the fact that they are an efficient and cost-effective means to collecting large amounts of data on a range of self-regulatory processes (Wolters et al., 2011). This is especially convenient given that self-regulation is a broad construct that encompasses many different processes. The data collected from survey instruments are easily quantifiable and can be quickly prepared for analyses using statistical software. Another advantage of self-report surveys is that the validity and reliability of the instruments can be determined. Once determined, the scales can be made available for researchers and/or teachers to use to diagnose student self-regulation deficiencies. As a result, many empirically validated instruments have been developed to assess self-regulation holistically and by individual process. Of the available self-regulation surveys, the most widely used Likert style instruments are the Learning
and Study Strategies Inventory (LASSI; Weinstein et al., 1987), and Motivated Strategies for Learning Questionnaire (MSLQ; Pintrich et al., 1993).

Although both the LASSI and MSLQ were framed around the social cognitive perspective of self-regulated learning, the scales measure self-regulated learning retrospectively, typically without any reference to a specific context. Perry and Winne (2006) argue that these traditional methods of measuring self-regulated learning assesses what students’ perceive their study habits to be, as opposed to the study habits they actually practice. In fact, Winne and Jamieson-Noel (2002) found evidence that students often report study habits that do not align with the study habits they actually put into practice, suggesting that students may not accurately calibrate their self-regulatory behaviors.

The LASSI, MSLQ, and other self-report instruments can be adapted to fit a specific context. For example, students may be instructed to report their study habits for a particular course. However, even when confined to the parameters of a course, the study habits practiced within the course can vary from task to task. Specifically, students’ study and learning habits vary according to the demands of a specific task within a course (Hadwin, Winne, Stockley, Nesbit, & Woszczyna, 2001). In fact, Hadwin et al. (2001) found that within one course, student employed different self-regulatory strategies for reviewing text, writing essays, and studying for an exam. Therefore, unless students are instructed to report on their study practices within a task-specific context, generalizations about overall study approaches can lead researchers to make inaccurate conclusions.
Finally, another limitation of surveys is that it assesses self-regulation as an aptitude or a trait (Winne & Perry, 2000). In surveys, the different self-regulatory processes are measured as aggregated traits that the student can exhibit “not very often at all” or “very often.” Although aggregation allows the researcher to statistically predict future behavior or achievement, peoples’ estimations of their tendencies are generally inaccurate, where they may under or overestimate the occurrence of different behaviors (Tourangeau, Rips, & Rasinski, 2000). Self-reported surveys also rely on the respondents’ memory of certain events or behaviors. Although information regarding the different behaviors should be easily accessed through memory, if information is not easily accessible then the responses may be distorted.

**Interviews.** During the late 1980’s Zimmerman and colleagues created a semi-structured interview, where students were asked how they would approach six hypothetical contexts. The questions and contexts were designed to measure different aspects of student self-regulation (Zimmerman, 2008a). The scale was called the Self-Regulated Learning Interview Scale (SRLIS: Zimmerman & Martinez-Pons, 1986, 1988), which consisted of open-ended response items with which the researchers would transcribe and code into 14 self-regulation categories. This interview also included Likert scale items that assessed students’ frequency of using the different strategies. According to Wolters et al., (2011) interviews can offer a range of options for researchers. Specifically, the questions can be specific or broad. The strength of the interview method is that it allows researchers to further question and probe student responses that are not clear or to delve deeper into in their responses. As compared to self-report surveys, where
student responses are limited to the constructs within the survey, interviews provide an avenue for students to further elaborate.

Although the interview method may provide a stronger depth of data regarding student self-regulation than the survey method, survey and interview methods share certain limitations. Specifically, both surveys and interviews are typically conducted outside of a natural setting and rely on students’ ability to recall and calibrate their self-regulation strategies, which prior research suggests is not very accurate (Winne & Jamieson-Noel, 2002). In terms of data collection, interviews inherently require much more coordination, time, manpower, and effort than surveys. As a result, sample sizes from interviews tend to be lower than sample sizes from survey research, which reduces the level of power in statistical analyses. Since qualitative data are collected during the interviewing process, a significant amount of time and effort must also be spent on transcribing and coding the data before quantitative analyses can be conducted. Finally, another limitation stems from the nature of qualitative methodology, which is that the biases of the interviewer are likely to influence student responses (Maxwell, 2005).

**Observations.** Observations involve directly observing or video recording students in their natural learning settings and coding their behaviors into appropriate self-regulation categories (Maxwell, 2005). Observations can range from very structured (e.g., including a checklist of behaviors observed) to very broad (e.g., open-ended observations, analyzing the data for emergent themes). One clear advantage of this approach is that self-regulation is assessed within a natural context that is often task-specific. As a result, data on student self-regulation behavior does not depend on student
recall or calibration—it is assessed as it occurs. This approach is often more fitting for populations that have trouble describing their behavior due to limited cognitive or linguistic abilities (Wolters et al., 2011).

Some of the disadvantages of this approach are that observations depend on observable actions. However, self-regulation is a combination of both cognitive (i.e., unobservable processes such as self-efficacy and metacognition) and behavioral (i.e., observable processes such as restructuring the environment and asking for help). To address these limitations, researchers can design a study with both interview and observation components (Perry, VandeKamp, Mercer, & Norby, 2002), where observed behaviors can be followed up upon during an interview to capture unobservable processes. Additionally, similar to interviews, observations require increased time and effort from the researchers to collect, prepare, and code the data.

**Think-alouds.** A commonly used method for assessing self-regulation as an event is through think-alouds (Azevedo, 2005; Greene & Azevedo, 2009). A think-aloud combines both interview and observation strategies to assess student self-regulation. Specifically, in a think-aloud, students are asked to verbally describe any thought processes and emotions while they are engaged in a learning task (Azevedo, 2005; Boekaerts & Corno, 2005). The sessions are typically recorded and coded for data analysis. Similar to observation strategies, think-alouds allow researchers to investigate self-regulation as it is occurring within a natural setting, thus not requiring students to recall or calibrate their self-regulatory behaviors.
A clear limitation to this approach is that it can be highly invasive to the students’ learning process. Cognitive thoughts are inherently internal and individuals may find it difficult to verbalize a process that is typically not verbalized (Boekaerts & Corno, 2005). This may be especially difficult for younger children or others with cognitive disabilities that may not have the skills to verbalize thought processes while engaged in a task. Another limitation is that the act of verbalizing thoughts itself influences the learning process (Greene, Robertson, & Costa, 2011). In fact, there is evidence to suggest that verbalizing cognitive thoughts while engaged in learning tasks is positively associated with achievement (Green et al., 2011).

Another limitation is that in many think-aloud studies, much data are gathered regarding students’ metacognitive processes and strategy use, however, students tend to be less vocal about their motivation (Bernacki, Aleven, & Nokes-Malach, 2014). Therefore, Bernacki et al. (2014) suggest that in order to investigate student motivational processes, a more explicit approach would be required, where researchers would need to prompt students to report their motivation during learning. Finally, similar to observations and interviews, collecting think-aloud data requires much time and effort as compared to survey methods.

**Diaries.** Diaries or logs require students to track, monitor, and record their learning experiences (Schmitz, Klug, & Schmidt, 2011). With proper structuring, the diary approach can measure several aspects of self-regulation within a meaningful context (Schmitz et al., 2011). For example, students can be directed to record their motivational feelings and goals after introduced to a task as well as their study and
monitoring strategies while engaged in the task. Afterwards, students can record their reactions and satisfaction to their outcome. According to Schmitz et al. (2011) the diary approach is designed to collect learning processes as they occur in real time or directly after learning, thus ecological validity with this approach is higher than traditional survey methods.

One major limitation to this approach is that it requires students to follow strict guidelines with completing diary entries. Oftentimes, students complete diary entries during their own time, not in the presence of researchers. Therefore, validity may be compromised if students do not follow the strict guidelines required from diary methodologies. Another limitation is similar to think-aloud methods, where the act of recording their learning processes in and of itself is beneficial to learning, which may confound any effects on achievement from self-regulation. Finally, the validity of diaries depend on both the motivation of the students to complete the entries and their writing skills (Spörer & Brunstein, 2006).

**Trace Methods.** Trace methods were developed by Winne and Perry (2000) to address the shortcomings of present methodologies for assessing self-regulation. Specifically, researchers suggest that the present survey and other qualitative approaches do not take into account the dynamic and context specific nature of self-regulation. They also suggest that assessment methods that attempt to capture the dynamic aspects of self-regulation tend to be rather invasive (e.g., think-alouds), where researchers often disrupt students while they are learning to question their thought processes or behaviors. Therefore, Winne and Perry (2000) developed learning software designed to indirectly
measure self-regulated learning through student actions. Specifically, the software includes a task that students are directed to engage in and within the task are options and tools that students can use to assist their learning process such as highlighting, hyperlinks, notes, and bookmarks. Depending on the frequency and length of time used on each tool, researchers are then able to infer the types of self-regulation processes students engage in during the learning process. Trace methods tend to be unobtrusive and do not rely on the ability for students to calibrate or recall self-regulation information. Instead trace methods are situated within an actual learning context measuring self-regulation processes in real time.

Some limitations with trace methods are that it is conducted on a computer and in a predesigned learning task. Although self-regulation is measured within an actual learning context, that online context may not be representative of the actual classroom environment and self-regulation strategies used through software may be different from the self-regulation strategies used during classroom learning. Therefore, Winne and Jamieson-Noel (2002) suggest that trace methods are the most accurate when utilized in conjunction with other methods of data collection.

**Microanalysis.** Similar to trace methods, another alternative approach that conceptualizes self-regulation as an event as opposed to a trait is microanalysis (Zimmerman, 2008a). Microanalysis is a method where one investigates students’ thoughts and actions as they are engaged in a meaningful, context specific task (Cleary, 2011; Cleary & Zimmerman, 2004; Kitsantas & Zimmerman, 2002; Zimmerman, 2008a). Microanalysis is a highly specific form of measurement where students’ perceptions and
behaviors are assessed in real-time at key points during learning. Typically, researchers collect data from students one-on-one while they are engaged in an actual learning task. Researchers would then ask students specific questions as they move through the phases of learning to directly measure each process. This form of measurement captures student self-regulation in real time, as it is occurring, resulting in high ecological and construct validity.

Although this method of measurement captures the dynamic and context specific nature of self-regulation, it can be invasive, time consuming, and requires a high amount of effort to collect data. Due to the one-on-one nature of this method, sample sizes also tend to be quite low, limiting the statistical power of quantitative analyses. Furthermore, teachers who want to diagnose their students’ self-regulation using this methodology would not be able to expend the time required with each student to capture their self-regulation deficiencies. Additionally, scales tend to be one item scales, preventing reliability coefficients from being calculated.

Overall, the available methodologies for measuring self-regulation seem to fall along a continuum of strengths and weaknesses by theory and practicality. On one end of the spectrum, the self-report survey method is efficient and cost effective, however it assumes that self-regulation is a global skill, lacks context, and relies on student ability to recall or calibrate their self-regulation. On the other end of the spectrum, microanalysis takes into account the dynamic aspect of self-regulation and questions student behaviors and cognitions as they occur, but is underutilized due to the extensive demands of collecting the data. In an attempt to bridge the gap between theoretical soundness and
practicality, this study designed a new approach to measuring self-regulation by designing a survey based on microanalytic methods. These next sections will describe the microanalytic methodology in more detail and introduce a new, modified microanalytical approach to measuring self-regulation.

**Microanalysis History and Background**

Bandura (1977) introduced the microanalysis methodology in a study that investigated the self-efficacy of individuals with severe snake phobias. In his study, individuals with snake phobias were asked to rate their self-efficacy to engage in various interactions increasing in difficulty with snakes. A model would demonstrate the activity with a snake while the participant observed. The participants were asked to rate their self-efficacy to engage in the observed tasks prior to witnessing the task, during observation of a model engaging in the task, after they attempted to engage in the task, and after they completed the task. Bandura found that different tasks with varying difficulty levels elicited different self-efficacy and behavioral outcomes. Bandura concluded that not only is self-efficacy a task specific concept, but global measures of self-efficacy are not as accurate predictors of performance as measures of self-efficacy at the individual task level.

Since this initial study, Bandura and other educational psychologists have advocated using the microanalytic methodology, specifically in studying self-efficacy and behavior modification. However, his work has been extended to self-regulation and learning (Cleary, 2011). In the late 1990’s Zimmerman and colleagues have further refined the approach to investigate how learners self-regulate athletic learning (Kitsantas
& Zimmerman, 2002; Kitsantas, Zimmerman, & Cleary, 2000; Zimmerman & Kitsantas, 1996, 1997). Recently, Zimmerman and colleagues have also used microanalytic methods to diagnose students’ self-regulation deficiencies in academic learning (DiBenedetto & Zimmerman, 2010, 2013; Cleary, Platten, & Nelson, 2008; Cleary & Zimmerman, 2004), however, only two studies to date have used the microanalytic method to comprehensively investigate the three phased model of self-regulation in the academic context (DiBenedetto & Zimmerman, 2010, 2013).

One of the first studies that used the microanalytic method to investigate self-regulation (Kitsantas & Zimmerman, 2002) assessed expert, non-expert, and novice volleyball players’ engagement in the three phased model of self-regulation. Students were individually observed as they practiced the volleyball serve and different questions assessing their self-regulation were asked at key points during the practice. For example, to assess students’ self-efficacy, the researcher asked the student to first rate her self-efficacy before she attempted the serve, and again after the student missed two serves consecutively. This procedure was followed for each of the different self-regulatory processes measured (12 processes total). Students were asked to respond to short, specific, closed and open-ended questions at key points of their practice. The questions asked were aligned with the corresponding phase that the self-regulatory process fell under (e.g., students were asked to delineate which strategies they planned to use before practicing the serve [forethought phase] and again while they were practicing [performance phase] and attributions for success or failure was asked at the end of the practice session [self-reflection phase]).
This microanalytic procedure allowed researchers to conclude that differences in skill level (expert, non-expert, and novice players) corresponded to differences in self-regulation. Specifically, the results showed that students’ self-regulatory practices successfully discriminated three different volleyball skill levels. Expert players reported the highest levels of adaptive self-regulation across all of the 12 measured processes than non-experts and novices. Additionally, microanalytic self-regulation methods revealed that the quality of engagement in self-regulation is more predictive of volleyball serving success above and beyond the amount of knowledge or experience in volleyball serving. Most notably, however, all of the scales combined together predicted 90% of the variance in volleyball serving skills. This finding not only has been replicated with findings from a basketball free throw practice task (Cleary & Zimmerman, 2001), but it also confirmed the reliability and predictive validity of microanalytic measures.

One of the first studies to use a comprehensive form of microanalysis to measure self-regulation in academic learning was DiBenedetto and Zimmerman (2010), where they investigated whether self-regulatory processes can differentiate between high, average or low achieving high school students in science. Students were instructed to read a science passage and respond to questions regarding the passage. As students were engaged in the task, researchers microanalytically assessed their self-regulated learning. For example, if the researcher observed a student highlighting a passage, the researcher would ask the student what he/she is doing and why. A total of 11 self-regulatory processes across the three phases were assessed. The results showed that indeed, students who achieved higher in science were more likely to adopt more adaptive self-regulation
processes than students who achieved lower in science, regardless of gender. In fact, girls achieved marginally higher than boys on the science test and reported higher self-satisfaction beliefs than boys. The authors concluded that besides self-satisfaction, boys were no different from girls on any of the other self-regulation processes measured. This may indicate that it is differences in self-regulation that lead to achievement differences, not gender.

**Critical Components of SRL Microanalysis**

Cleary (2011) outlines four components of microanalysis in self-regulation: 1) individualized assessment, 2) multiple self-regulation processes, 3) question structure, 4) correspondence between self-regulation questions and three phased model, and 5) scoring and coding. The sections below describe the characteristics of each component.

**Individualized assessments.** In terms of individualized assessments, Cleary (2011) emphasized that in order to reduce social influences, students should be administered the questions one-on-one with the researcher. However, microanalytic procedures have also been employed to students in small groups, where students responded to questions on a piece of paper while they engaged in a task (Cleary et al., 2008; DiBenedetto & Zimmerman, 2010). For example, DiBenedetto and Zimmerman (2010) designed an SRL microanalytic assessment for students to complete as they completed a reading task. Questions regarding specific self-regulatory processes were placed strategically following the three phased self-regulation model. For example, to assess students’ metacognitive monitoring, questions such as, “How confident do you feel about your answer to question 1” were placed after each test question the students
answered (e.g., performance phase process) and satisfaction was assessed immediately after students received the grade to their reading test (e.g., self-reflection phase process).

This present study deviated from traditional microanalysis methods by microanalytically diagnosing self-regulation as a class as opposed to individual or small group assessment. This was done to take into account the reality of large college classrooms and research data collection procedures. Professors are not likely to have enough time to diagnose individual student self-regulation. Therefore, having a diagnostic tool to merge and administer with a classroom test would greatly reduce the amount of time and effort required of professors to examine their students’ self-regulation and motivation.

**Multiple self-regulation processes.** The second important component of SRL microanalysis is the assessment of multiple self-regulation processes across the three phases. One of the main goals of SRL microanalysis is to capture self-regulation holistically as opposed to its individual components. Self-regulation, after all, is characterized by the interaction of the three phases and the processes within them. Therefore, the more processes within the different phases assessed, the more valid and precise the measurement is.

**Question structure.** The third component of SRL microanalysis is structuring the microanalytic questions so that they are brief, straightforward, directly related to the task at hand, and based off the operational definition of the process (Cleary, 2011; Zimmerman, 2000). For example, if the SRL microanalysis was designed to assess students’ math performance, a question regarding their self-efficacy would be “How
confident are you that you will answer this question correctly?” This question is brief, directly related to the task, and is taken from the definition of self-efficacy (e.g., the degree to which an individual feels confident that he/she can successfully accomplish a goal/task [Bandura, 1977]).

Microanalytic questions can also be open or closed ended. Students respond on a Likert scale for close ended responses. Closed ended questions are generally those that assess self-motivational beliefs from the forethought phase (e.g., self-efficacy, task value) and those that do not require the student to indicate their behaviors or beliefs. However, other self-regulatory behavioral processes should be open-ended (e.g., causal attributions, self-evaluation). For example, traditional methods of assessing task strategies involve asking students the degree to which they engage in elaboration, summarization, etc., in an aggregated form. However, microanalytic procedures allow the student to specifically delineate the different strategies that he/she actually plan to use or are using. Scoring would then involve counting the number of strategies that the student used or the coding the quality of the strategy (DiBenedetto & Zimmerman, 2010, 2013).

**Correspondence between self-regulation questions and three phased model.**

The timing to which specific processes are assessed must correspond to whichever phase of learning the student is engaged in. For example, when assessing a student’s metacognitive monitoring, the question should be asked while the student is performing the task (e.g., performance phase “during the task”). Therefore, forethought questions should be administered before the student engages in the learning task, performance questions should be administered while the student is engaged in the task, and self-
reflection questions should be administered after the student has completed the task. This type of real-time assessment of self-regulation is key to the validity of the microanalytic method.

**Scoring and coding responses.** Scoring of closed ended, Likert type responses are generally straightforward and follow the traditional survey methods. Generally, in microanalytic methodology, many of the processes are assessed with one item scales. In terms of open-ended items, coding follows a prescribed manual and rubric based on prior research. For example, for causal attributions, students asked why they achieved that particular grade on their math test. The coding procedures will follow Weiner’s (1986) classification of attributions along the three major dimensions (e.g., stability, controllability, locus of control). Coded open ended responses thus must be grounded in a strong theoretical framework relevant to the specific process (e.g., attributions follow Weiner’s attributions theory).

**Validity of SRL Microanalysis**

Researchers have highlighted some potential validity pitfalls of the microanalytic method which include: 1) low sample sizes and 2) one item subscales (Kitsantas & Zimmerman, 2002). Because this methodology is highly contextualized, task specific, and requires researchers and instructors to conduct personalized one-on-one interviews, sample sizes are generally low. Kitsantas and Zimmerman (2002) suggest that although low sample size may compromise statistical power, the context-specific nature of this methodology actually make the questions less ambiguous and thus more predictive of performance than traditional survey items.
One item subscales may also compromise the validity and reliability of the methodology. However, the strength of the microanalytic method lies with its highly contextualized, real-time, theoretically based nature. Kitsantas and Zimmerman’s (2002) study found that out of the 12 different self-regulatory variables measured microanalytically, reliability estimates of all the single item scales combined under one scale was .90 and Kappa coefficients ranged from .81 to .98 for open-ended coded responses. In terms of validity, content validity is established by grounding the microanalytic questions on strong theoretical models (e.g., three phased self-regulation model) and basing the questions on operational definitions. There is also evidence that microanalytic methods have strong predictive validity. Specifically, Kitsantas and Zimmerman (2002) merged several microanalytic self-regulation scales into one composite self-regulation measure to predict performance. This composite self-regulation score explained 90% of the variance in volleyball serving skills, which was the ultimate outcome of interest in this particular study. Further, Kitsantas and Zimmerman (2002), Cleary and Zimmerman (2001), and DiBenedetto and Zimmerman (2010) established differential validity by showing that differences in physical and academic performance can be explained by differences in self-regulated learning.

More recently, DiBenedetto and Zimmerman (2013) investigated the construct and predictive validity of the microanalytic approach in the learning context. A total of 51 11th grade students participated in the study. Student microanalytically measured self-regulation was compared against teacher ratings of student self-regulation to establish construct validity. Predictive validity was assessed by comparing the variance accounted
for between student microanalytic measures and teacher rating scales on two types of student achievement measures: a 10 question, short answer tornado test and a conceptual model test (e.g., drawing out and explaining how a tornado forms). In terms of construct validity, all of the microanalytic scales correlated positively with the global self-regulation teacher rating scale, indicating an acceptable level of construct validity. In terms of predictive validity, the results showed that all of the microanalytic measures altogether accounted for 50% of the variance in student achievement on the tornado test, and the teacher rating scale added at the end of the model contributed an additional 3% of variance. In terms of the conceptual model test, the self-regulation variables all together explained 81% of the variance in student achievement, and the teacher rating scale added at the end of the model did not contribute a significant amount of variance above and beyond the self-regulation variables. Overall, this study concluded that microanalytic measures of student self-regulation have strong construct and predictive validity.

**Self-Regulation Assessment Tool: A Modified Microanalytical Approach**

This study will use a form of microanalysis, incorporating all of the important components of microanalysis as described by Cleary (2011) (i.e., multiple self-regulation processes, question structure, correspondence between self-regulation questions and three phased model, and scoring and coding) except for individualized assessment. Specifically, this approach will be administered as a self-report survey to a classroom of students situated in a meaningful context where self-regulation processes are assessed as it occurs in real time. This approach attempted to merge the theoretical advantages of a classic microanalytic methodology with the practical advantages of a self-report survey.
There was one major drawback for not incorporating individualized assessment which was the inability to observe behaviors and ask questions during specific moments of the self-regulation process. For example, during individualized assessment, the observer may ask the student why he/she underlined a word in a test question and record it as a test taking strategy. In whole group assessment, that level of behavior is not probed and the accuracy of the assessment would rely on student self-report. However, the drawbacks of student self-report (e.g., inaccurate calibration or recall) are minimized because: 1) students are asked to report their self-regulation processes as they are actually engaged in the processes, and 2) students are asked to report their self-regulation processes immediately after they have engaged in them. Ultimately, the time between self-regulation engagement and self-regulation reporting is significantly shortened, making the ecological validity of this approach much stronger than in traditional self-report surveys. Finally, the last major advantage to administering a microanalysis to a group is that it is more convenient and more practical for researchers and college professors who want to investigate their students’ level of self-regulation and motivation. Therefore, this study did not utilize the classic microanalytic method, but a modified form of the method that is more conducive to large group, whole class administration.

To summarize the literature up to this point, it can be concluded that:

1. Zimmerman’s three phased model of self-regulation is a powerful and comprehensive construct that centers on students’ self-motivated and self-directed role in their own learning process. Research has confirmed that differences in achievement are largely due to differences in self-regulation processes (Cleary &
2. Stereotype threat is a social phenomenon that causes individuals of a stigmatized social group to underperform when exposed to a negative stereotype (Steele & Aronson, 1995). Unfortunately, researchers have yet to uncover practical mediators that would allow students and teachers to combat this effect (Good et al., 2008; Nguyen & Ryan, 2008; Schmader, 2002; Stoet & Geary, 2012).

3. The most popular assessment strategies for self-regulation are self-report measures, however, this method inherently assumes that self-regulation is a global and static skill when self-regulation theoretically changes across time and events (Cleary, 2011; Hadwin et al., 2001; Winne et al., 2001; Winne & Perry, 2000). Methodologies that take into account the dynamic nature of self-regulation, like microanalysis, are time-consuming and costly, making the research process rather inefficient. Therefore, a modified microanalytic approach was developed to be administered as a self-report survey in an attempt to bridge the gap between practicality and theoretical soundness.

Now, to merge all three components of the literature together, these next sections will review research that provide insight into the relations between self-regulation and stereotype threat, and why this proposed modified microanalytic methodology would be the most ideal approach to investigating these relations.
The Relationship between Self-Regulation and Stereotype Threat

The connection between self-regulation and stereotype threat may be obscured because stereotype threat is an interdisciplinary concept. Specifically, there are several terminologies unique to educational and social psychology, but essentially hold the same meaning such as task value or domain identification and metacognition or doubt or optimism. The review below attempted to identify evidence in current literature that merge two broad disciplines: self-regulation and stereotype threat.

**Forethought phase: Self-efficacy.** Although researchers have directly examined the relationship between self-efficacy and stereotypes, the results are inconsistent. For example, Spencer et al. (1999) hypothesized that women’s math performance in low stereotype threat conditions may not have increased because of the absence of stereotype threat, but by the elevation of performance expectations, or, self-efficacy. In other words, differences in self-efficacy may explain performance differences as opposed to high or low stereotype threat. However, the results showed that students’ level of self-efficacy did not mediate the differences in achievement across stereotype threat conditions. Conversely, Hollis-Sawyer and Sawyer (2008) argue that low stereotype threat may cause an increase in self-efficacy. In their study, Hollis-Sawyer and Sawyer (2008) found that individuals exposed to high stereotype threat experience elevated levels of anxiety (which translates to lowered levels of self-efficacy), leading them to conclude that test anxiety/self-efficacy may be the mechanism underlying stereotype threat. Furthermore, Aronson and Inzlicht (2004) found that African Americans who were more vulnerable to stereotypes reported stronger fluctuations of self-efficacy than African Americans who
were less vulnerable to stereotypes. Some of these findings suggest that individuals who are more likely to be affected by stereotype threat experience a less stable and adaptive sense of efficacy.

**Forethought phase: Attainment task value.** In terms of attainment task value, research has found that the more importance stigmatized students place on doing well in a certain task, the more vulnerable they are to the effects of stereotype threat (Aronson & Good, 2002; Aronson et al., 1999). Specifically, Aronson and Good (2002) found that stigmatized students who did not value doing well in a task under stereotype threat conditions had significantly performed higher than students who reported higher levels of value in a low stereotype threat condition. The authors rationalized that the anxiety induced by stereotype threat was more pronounced in those who actually wanted to do well, whereas students who did not care about doing well had less of a reason to be impacted by the anxiety.

The term used in social psychology to describe value in a particular content area is domain identification, which similar to task value in educational psychology. Previous research in stereotype threat has confirmed that women who have less domain identification (or task value) with math, the less pressure they feel to perform (Lesko & Corpus, 2006). This decreased pressure would result in a lower likelihood to experience stereotype threat. The assumption is that students who have low domain identification with math are less motivationally invested and failure would not threaten a vested aspect of their identity.
For example, Lesko and Corpus (2006) hypothesized that women who are highly math identified (i.e., thereby having high math task value) will be more likely to discount the form of evaluation (i.e., discounting is defined as rejecting the validity of an evaluation) that is linked to stereotypes than women who are less math identified. The authors rationalized that women who are highly math identified cannot "disidentify" with math in the face of stereotype threat, given that their major requirements and future career goals will likely require them to value and apply their math skills. Therefore, an ego-protective strategy that highly math identified women may use would be to discount the evaluation that threatens their sense of ability in math. The results not only reconfirmed the presence of the stereotype threat effect in women’s math performance, but it also showed that math identification was a significant moderator of the relationship between discounting and stereotype threat. In other words, women who were highly identified with math and subjected to stereotype threat reported higher levels of discounting than any other group. Additionally, although the authors hypothesized that discounting plays an ego-protective role, posttest analysis did not reveal that highly math identified females who discounted had higher or lower self-esteem reports than low math identified females.

Therefore, according to social psychology research, stereotype threat only occurs when a negative stereotype threatens an individual’s self-definition, because high identifiers are more likely to reach a higher level of stimulation when faced with a negative stereotype than low identifiers. However, this relationship does not align with social cognitive theory and self-regulated learning. Specifically, students who value a
task (or with high domain identification) would be more motivated to engage in self-regulatory strategies, thus increasing the likelihood of academic success (Wigfield & Eccles, 2000). This relationship is reversed in stereotype threat research, where stigmatized students who value the subject matter are more susceptible to stereotype threat effects.

Both explanations from social and educational psychology are plausible and have been supported. Students who highly value a task or a domain are more likely to achieve higher than students who do not highly value a task or a domain, but students who highly value a task are also more likely to be negatively impacted by a threatening stereotype that undermines their social group. In terms of practicality, however, this cannot be used to combat stereotype threat effects, because teachers must encourage girls and women to value masculine fields such as math and science, not to avoid, devalue, or disidentify with these fields. This complex relationship and opposing results regarding task value may be better understood if it was investigated not as an individual construct, but as a process within self-regulation.

**Performance phase: Attention focusing strategies.** One important mechanism that is thought to underlie stereotype threat in social psychology is cognitive load and deficits in working memory (Aronson & Steele, 2005; Schmader & Forbes, 2008). Researchers have found that a large amount of cognitive resources are consumed when students are faced with a negative stereotype regarding their achievement in a domain that they value and identify with. Those cognitive resources that should be applied towards completing the task are shifted towards the threat, thereby limiting the efficiency
of working memory (Schmader & Johns, 2003). This cognitive burden that stigmatized students face has been documented in previous research (Croizet et al., 2004; Inzlicht, McKay, & Aronson, 2006; Schmader & Johns, 2003), where stereotype threat can be explained by deficiencies in cognitive resources.

In terms of women and math specifically, Schmader and Johns (2003) found that women under high stereotype threat conditions not only performed lower on a math test than women under low stereotype threat conditions, but also recalled fewer words from their working memory. Specifically, Schmader and Johns (2003) hypothesized that targeted individuals under stereotype threat conditions would underperform on cognitive tasks due to deficits in working memory. By making a negative stereotype regarding women’s math performance salient right before a math test, their working memory capacity was hindered, ultimately suggesting that working memory mediates the stereotype threat effect.

To test this hypothesis, Schmader and Johns (2003) completed several experiments. The first experiment confirmed that women under stereotype threat conditions had indeed exhibited a decrease in working memory as compared to women not under stereotype threat conditions. The second experiment extended the results of the first experiment, where it showed that stereotypes made implicitly salient to Latinos (e.g., test measuring intelligence with no explicit statement of Latinos, but students were instructed to indicate their ethnic background on the test) caused a reduction in their cognitive capacity. Finally, the third experiment tested mediation effects, where working memory capacity would fully mediate stereotype threat on women's math performance. A
series of regressions ultimately revealed that stereotype threat did not significantly predict math test performance when working memory capacity was entered as the predictor. These results all confirm that one of the underlying causes of the stereotype threat effect is limited working memory capacity, suggesting that working memory is an important underlying factor in stereotype threat.

Johns, Inzlicht, and Schmader (2008) further investigated these initial findings. Specifically, Johns et al. (2008) found that women under high stereotype threat conditions have impaired executive functioning due to their attempts to regulate their emotional reactions to the threat, resulting in lowered math performance as compared to women in low stereotype threat conditions. Therefore, students who attempt to regulate their emotions during test taking experience a depletion of cognitive resources that are necessary to effectively complete the task. However, Johns et al. (2008) also found that stereotype threat effects were diminished when women were provided coping mechanisms through reappraising the anxiety. Specifically, after students were exposed to the stereotype threat and before they engaged in the task, students were told that research shows that the anxiety that they may experience does not affect, but actually facilitate performance on that particular test.

Schmader and Johns (2003) and Johns et al. (2008) provided converging evidence that working memory may indeed be one of the root causes of stereotype threat. Although these studies provide information regarding the rudimentary processes underlying stereotype threat, one of the major limitations of this study was that it was conducted in a laboratory and unauthentic setting. However, these studies do provide a strong
framework for informing this present study. Specifically, a powerful aspect underlying stereotype threat is the anxiety that it induces and how individuals control or regulate that anxiety under stereotype threat conditions. Therefore, refocusing the attention from regulating the anxiety to the specific task may be an important aspect of self-regulation under stereotypic conditions. Attention focusing strategies allows students to improve their attention to and concentration on the given task (Boekaerts & Niemivirta, 2000; Corno, 1993; Zimmerman, 2008a). Therefore, it would be important to assess students’ ability to focus their attention under stereotype threat conditions.

**Performance phase: Metacognitive monitoring.** The relationship between metacognition and stereotype threat has yet to be thoroughly investigated by educational or social psychology researchers. Metacognition, as previously described, is a process by where individuals plan, monitor, and evaluate progress toward a given goal. In educational psychology, metacognitive monitoring more specifically refers to how students monitor their metacognitive processes. For example, students may ask themselves whether they understand a certain question or whether the answer they provided is complete or correct. Social psychologists are more interested in the metacognition towards the situation, whereas educational psychologists are more interested in the metacognition towards the academic task. Specifically, Johns and Schmader (2010) suggest that stigmatized students do not passively cope with the stereotypes that they face but actively attempt to respond to the situation that they are placed in. When attempting to cope with an ego-threatening situation, Johns and Schmader (2010) suggest that stigmatized students may experience heightened arousal and
become more aware, or metacognitive, about their own behavior, performance, and reaction of others. This type of awareness may take a toll on the cognitive and executive resources required for successful completion of a difficult task.

For example, Schmader, Forbes, Zhang, and Mendes (2009) examined whether confident or doubtful judgments (i.e., metacognitive judgments) about performance would inhibit working memory function in a sample of college students completing a working memory test. Three studies were conducted. In the first study, students were primed with either confident or doubtful judgments while completing a working memory test. Working memory was assessed through the number of neutral words students were able to memorize. The neutral words were presented through a flashing computer screen that alternated between the neutral words and different sentences. To prime confidence or doubtfulness, the sentences were either confident in nature (e.g., I am confident that my team will win big tonight) or doubtful in nature (e.g., I am doubtful that my team will win big tonight). Working memory was assessed through presenting students with a list of neutral words presented in a computer screen. The results showed that in addition to a stereotype threat effect, minority students primed with doubt exhibited lower working memory abilities than minority students primed with confidence.

In the second study, the authors attempted to replicate the ethnicity effects with gender. The same working memory test to investigate “intelligence” was used in this study. However, in order to evoke stereotype threat, the authors informed the women that they would be completing a difficult math test to investigate gender differences in mathematical ability. The results of this study showed that confidence and doubt
judgments moderated the relationship between anxiety and working memory in stereotype threat conditions. Specifically, women with high levels of anxiety experienced deficiencies in working memory when primed with doubt. However, women with high levels of anxiety did not experience deficiencies in working memory when primed with confidence.

The third and final study was conducted to investigate the above results in more naturalistic settings. Women’s mathematic performance was investigated and self-doubt and situational reappraisal (defined as reinterpreting the stressful situation to make it less stressful) were measured with questionnaires. Anxiety was measured with levels of epinephrine and norepinephrine present in students’ saliva. The results showed that women in stereotype threat conditions who had higher levels of anxiety tended to adaptively reappraise the situation thus were more likely to achieve higher on the math test than women with high anxiety who did not tend to reappraise the situation. Ultimately, Schmader et al. (2009) concluded that students who have a more positive mindset are more able to overcome stereotype threat effects.

Only one empirical research study was found that investigated metacognition and stereotype threat, which highlights the need for further, more naturalistic and holistic studies. Although thoughts of doubt and optimism are not exactly how educational psychologists conceptualize metacognition, it is related. Specifically, metacognitive monitoring is a broad term used to describe how students monitor their understanding. It can be argued that metacognitive students do express skepticism or optimism in their answers on a given test—however, whereas social psychologists consider this very
skepticism or optimism as metacognition, educational psychologists further include
aspects such as whether the student actually revisits the material (e.g., look through notes
to confirm answers, rereading passages/questions). Therefore, in terms of self-regulation,
metacognition is part of a larger process.

**Self-reflection phase: Causal attributions.** In educational psychology,
attributions toward a given outcome of a task are investigated. For example, a student
may attribute high marks on her test to the high amount of studying prior to taking the
test. By contrast, in social psychology, attributions toward the affect are commonly
investigated. For example, a woman may attribute her high anxiety on a math test to her
lack of studying the night before. Therefore, educational and social psychology
approaches attributes from different perspectives.

As previously discussed, attributions for success and failure play an important
role in informing how future tasks should be approached (Schunk et al., 2008). Generally,
students who make internal and controllable attributions (e.g., effort) typically attain
more adaptive academic outcomes than students who make external and uncontrollable
outcomes (e.g., luck). One study has directly investigated whether the types of
attributions females make may differ according to stereotype threat conditions.
Specifically, Koch, Muller, and Sieverding (2008) found that women in a high stereotype
threat condition had attributed failure to more internal explanations (low ability) and
attributed men’s failure to more external outcomes (technical problems with equipment)
than women in a low stereotype threat condition. Similarly, Schmader et al. (2004) found
that under high stereotype threat conditions, women were more critical of their
performance on a science test than men, even when the women and men achieved at equal levels. However, research has traditionally found that women’s criticism towards their own performance in masculine domains is unfortunately rather typical. Specifically, research has shown that women tend to be more critical of their own abilities and make more maladaptive attributions to their performance in male-dominated domains (e.g., science and math) than men (Dickhäuser & Stiensmeier-Pelster, 2002; Hyde, Fennema, Ryan, Frost, & Hopp, 1990).

As reviewed above, there is evidence that several self-regulation processes may contribute to explaining stereotype threat (i.e., self-efficacy, attainment task value, attention focusing strategies, metacognitive monitoring, and causal attributions). Most of the stereotype threat research was conducted in highly controlled laboratory studies, highlighting a need for researchers to investigate stereotype threat as it occurs in the natural classroom setting (Cromley, Perez, Wills, Tanaka, Horva, & Agbenyega, 2013, Good et al., 2008; Nguyen & Ryan, 2008). Like self-regulation, stereotype threat is a dynamic construct and its effect on achievement can vary across events and individual factors such as self-efficacy (Aronson & Inzlicht, 2004), math value or domain identification (Aronson & Good, 2002), or working memory (Aronson & Steel, 2005; Schmader & Forbes, 2008). Therefore, it would be important to measure self-regulation with methods designed to take into account the natural context that it occurs in such as microanalysis. In this study, a modified form of microanalysis was used to assess self-regulation processes as it occurs under stereotype threat conditions in a naturalistic classroom setting. The major goal of this study was to first investigate the initial
predictive validity of this modified approach, followed by exploring how this approach
can identify the nuances in self-regulation under stereotype threat conditions.

**Research Questions and Hypotheses**

The first research question along with the sub-questions all attempt to investigate
the initial predictive validity evidence of the SRAT:MMA. The second research question
attempted to investigate whether the SRAT:MMA can identify the self-regulation factors
that alleviate stereotype threat effects on women in math.

1. Can the SRAT:MMA diagnose college student self-regulated learning at the
classroom level during test taking in a math course?
   a. Do students with high self-regulation perform better than students with
      low self-regulation on a math test? It was hypothesized that students with
      low self-regulation will indeed perform lower than students with high self-
      regulation on a math test.
   b. Does the relationship between the different phases of self-regulation vary
      across levels of student achievement on a math test? It was hypothesized
      that the relationship between the different phases of self-regulation will
      vary across levels of student achievement, where students who performed
      high on the math test will exhibit more positive patterns of self-regulation
      than students who performed low on the math test.
   c. Can the self-regulation variables, as measured by the SRAT:MMA,
      predict college student achievement on a math test? It was hypothesized
that the self-regulation variables as measured by the SRAT:MMA can predict college student achievement on a math test.

2. Can self-regulation processes, as measured by the SRAT:MMA, alleviate stereotype threat effects in women enrolled in undergraduate math courses? It was hypothesized that the SRAT:MMA will be able to identify the self-regulatory factors that alleviate stereotype threat effects in college women.
Chapter Three

The major purpose of this study was to investigate the initial validity of the SRAT:MMA, which attempted to modify the full microanalytic approach to be administered as a self-report survey. Although stereotype threat and self-regulation has been somewhat investigated within the social and educational psychology fields, researchers have yet to combine these perspectives comprehensively to explain stereotype threat. Therefore, the secondary purpose of this study was to investigate whether self-regulation, as measured by SRAT:MMA, can alleviate stereotype threat effects. This following chapter will describe the methodology that was undertaken to investigate these research purposes.

Study Design

This study occurred in two phases. Phase One was conceptual and included a process for survey development. Phase Two involved administering the instrument to study participants under high and low stereotype threat conditions. These next sections will describe Phase One and Phase Two of the study.

Phase One: SRAT:MMA Development

Items for the SRAT:MMA were initially generated from a review of past literature. If items were not available from past microanalytic studies, then the alternative option was to adapt relevant items from validated survey instruments. If items from
validated survey instruments were not available then the last option was to create the items based on the operational definition of what the item was intended to measure (Cleary, 2011).

Two strategies were used to identify past microanalytic studies which included: searching through the PsycInfo and Education Resources Information Center (ERIC) databases for the keywords “microanalysis” or “microanalytic” and “self-regulation” and searching for articles that have cited handbook chapters or theoretical articles regarding self-regulation microanalysis (Cleary, 2011; Zimmerman, 2008a). Only prior studies that investigated the relationship between self-regulation and academic learning were considered. Out of the all the articles located, two were full microanalyses that examined a comprehensive form of self-regulation (DiBenedetto & Zimmerman, 2010, 2013), one theoretical article that provided specific microanalytic questions that can be used when diagnosing student self-regulation deficiencies (Cleary & Zimmerman, 2004), and two empirical articles that investigated a portion of self-regulation using microanalytic items (Chen, 2002 [self-evaluation microanalytic question]; Cleary et al., 2008 [attributions microanalytic question]). The full microanalytic studies (DiBenedetto & Zimmerman, 2010, 2013) were both designed and formatted to assess self-regulation from individual students. For example, the causal attribution item originally from DiBenedetto and Zimmerman (2010) was: “Why do you think you didn’t do better on this particular test question on tornado development? Please explain.” This item would be framed this way if the student did not do well on the test. If the student did very well on the test, the question would be framed: “Why do you think you did so well on this particular test
question on tornado development? Please explain.” Therefore, to adapt the item for a self-report group survey, the item was reworded to: “Please describe why you think you received this particular grade on this math test.”

If a microanalytic item was not available from past research, the second option was to adapt an item from a validated survey instrument. As previously discussed, most items from validated survey instruments measure self-regulation as a global construct. Therefore, all global survey items taken from validated instruments were adapted to take into account this present context, which was self-regulation processes used during a math test. If the item was not available from prior studies or validated survey instruments, then it was created based on the microanalytic guidelines from Cleary (2011). Specifically, the items would need to be context specific, short, direct, and based on the operational definition of the self-regulation process. The purpose of these procedures was to ensure the content validity of the assessment. In other words, all of the items must directly assess the operational definition of each self-regulation process to ensure that the items captured what it was intended to capture.

The initial draft of items was submitted to an expert in microanalysis and self-regulation for review. After some initial changes, the survey was then administered and reviewed by group of 12 (8 males and 4 females) college students ages 20-22. Students were asked to complete the survey in its entirety, which had also included three math questions from an online practice SAT test to mimic the actual survey as much as possible.
After determining the amount of time it took for students to complete the survey, they were asked to participate in a focus group to provide any feedback regarding the clarity of the survey. Based on the feedback and the responses, the self-evaluation question was revised. Specifically, students’ answer to the original question from Cleary and Zimmerman (2004) (“How do you determine whether you did well on your math test?”) was rather ambiguous, with responses such as “if I got a good grade,” or “anything above a B is a good grade.” Therefore, to further refine the question to measure what the self-evaluation item was intended to measure, the question was changed to a multiple choice format, where students specifically selected whether they evaluated their performance based on their personal goal, prior performance, or social comparison. See the guidelines used to develop the survey questions in Table 1.
<table>
<thead>
<tr>
<th>Self-Regulation Processes</th>
<th>Operational Definition</th>
<th>Microanalytic Item Available from Past Research?</th>
<th>Item Adapted for Current Study</th>
<th>Scale of Measurement</th>
<th>Coding Scheme</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forethought</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Goal Setting: Grade Goal</td>
<td>The goal that students want to achieve on the math test.</td>
<td>Yes: “Do you have a goal you are trying to achieve on your math test? Explain.” (Cleary &amp; Zimmerman, 2004)</td>
<td>What is the grade that you want to achieve on the math test you are about to take? Out of five total questions:</td>
<td>0-5 (The raw score/goal that students wanted to achieve on the math test)</td>
<td></td>
</tr>
<tr>
<td>Goal Setting: Goal Difficulty</td>
<td>The degree of difficulty it will be to achieve the goal. Process goal: Goal that focuses on following the specific processes of solving the math problem Outcome Goal: Goal that focuses on getting the math question right.</td>
<td>No</td>
<td>How difficult will it be for you to reach this goal on this test? Is it more important for you to focus on: a) the processes of solving the problems or b) getting the answers right</td>
<td>1-7 Likert scale</td>
<td>1 = Not very difficult 7 = Very difficult</td>
</tr>
<tr>
<td>Goal Setting: Process or Outcome Goal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Multiple choice 1 = Process goals 0 = Outcome goals</td>
</tr>
<tr>
<td>Self-Efficacy</td>
<td>Confidence in one’s ability to achieve the goal.</td>
<td>Yes: How confident do you feel in your capability to learn and remember all of the material on tornados for this passage?” (DiBenedetto &amp; Zimmerman, 2010)</td>
<td>On a scale of 1-7, how confident are you that you will be able to reach the grade you want to achieve on this math test?</td>
<td>1-7 Likert scale</td>
<td>1 = Not very confident 7 = Very confident</td>
</tr>
</tbody>
</table>

(Table continues)
<table>
<thead>
<tr>
<th><strong>Self-Regulation Processes</strong></th>
<th><strong>Operational Definition</strong></th>
<th><strong>Microanalytic Item Available from Past Research?</strong></th>
<th><strong>Item Adapted for Current Study</strong></th>
<th><strong>Scale of Measurement</strong></th>
<th><strong>Coding Scheme</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Strategic Planning</td>
<td>The test taking or math strategies that students plan on using during the math test.</td>
<td>Yes: “Do you have any particular plans for how to read this passage and take the test?” (DiBenedetto &amp; Zimmerman, 2010)</td>
<td>What strategies are you planning on using to help you succeed on this test? For example, any test taking strategies techniques, methods, or line of attacks?</td>
<td>Yes/No</td>
<td>1 = Yes 0 = No</td>
</tr>
<tr>
<td>Forethought</td>
<td></td>
<td></td>
<td>If yes, what strategies do you plan on using?</td>
<td>Open-ended</td>
<td>Count the number of testing strategies reported.</td>
</tr>
<tr>
<td>Task-Specific Task Value</td>
<td>The level of importance to do well on this specific math test.</td>
<td>No: From the Self- and Task-Perception Questionnaire (Eccles &amp; Wigfield, 1995): “How important is it for you to get good grades in math?”</td>
<td>How important is it for you to do well on this math test?</td>
<td>1-7 Likert Scale</td>
<td>1 = Not very important 7 = Very important</td>
</tr>
<tr>
<td>Domain-Specific Task Value</td>
<td>The level of importance to do well in math in general.</td>
<td>No: From the Self- and Task-Perception Questionnaire (Eccles &amp; Wigfield, 1995): “I feel that, to me, being good at solving problems which involve math or reasoning mathematically is… (not at all important, very important)</td>
<td>How important is it for you to be good at math?</td>
<td>1-7 Likert Scale</td>
<td>1 = Not very important 7 = Very important</td>
</tr>
</tbody>
</table>

*(table continues)*
<table>
<thead>
<tr>
<th>Self-Regulation Processes</th>
<th>Operational Definition</th>
<th>Microanalytic Item Available from Past Research?</th>
<th>Item Adapted for Current Study</th>
<th>Scale of Measurement</th>
<th>Coding Scheme</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Performance</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metacognitive Monitoring:</td>
<td>The extent to which</td>
<td>No</td>
<td>How confident are you</td>
<td>1-7 Likert</td>
<td>1 = Not very</td>
</tr>
<tr>
<td>Understanding of Question</td>
<td>students are aware of</td>
<td>their comprehension while performing a task.</td>
<td>about your understanding of</td>
<td></td>
<td>confident</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>this question?</td>
<td></td>
<td>7 = Very</td>
</tr>
<tr>
<td>Metacognitive Monitoring:</td>
<td>The extent to which</td>
<td>Yes: “How confident do you feel about your</td>
<td>How confident are you</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Confidence in their</td>
<td>students are aware of</td>
<td>answer to question X?” (DiBenedetto &amp;</td>
<td>that you have answered this</td>
<td></td>
<td></td>
</tr>
<tr>
<td>answers</td>
<td>their progress while</td>
<td>Zimmerman, 2010)</td>
<td>question correctly and</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>performing a task.</td>
<td></td>
<td>completely?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attention Focusing</td>
<td>Strategies used to</td>
<td>Yes: What do you do when you don’t feel like</td>
<td>At any point during the test,</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strategies</td>
<td>maintain focus on the</td>
<td>studying for your math test? (Cleary &amp;</td>
<td>did your concentration</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>test.</td>
<td>Zimmerman, 2004)</td>
<td>wander?</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>If yes, what did you do</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>to refocus your concentration?</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Open-ended</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Count number of attention</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>focusing strategies.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*(table continues)*
<table>
<thead>
<tr>
<th>Self-Regulation Processes</th>
<th>Operational Definition</th>
<th>Microanalytic Item Available from Past Research?</th>
<th>Item Adapted for Current Study</th>
<th>Scale of Measurement</th>
<th>Coding Scheme</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-Evaluation</td>
<td>How students evaluate their performance on the math test with either personal mastery criteria, past performance, or social comparison.</td>
<td>Yes: “How do you determine if you performed well on your math test?” (Cleary &amp; Zimmerman, 2004)</td>
<td>Please complete the following sentence: I feel that my performance on this test was successful/unsuccessful because (Circle a, b, or c below). a) I did/didn’t do better than the class average b) I did/didn’t do better than my previous tests c) I did/didn’t meet my goal</td>
<td>Multiple choice</td>
<td>1 = Social comparison* 2 = Prior performance 3 = Personal mastery criteria</td>
</tr>
<tr>
<td>Causal Attributions</td>
<td>How students explain their performance on the math test.</td>
<td>Yes: “Why do you think you didn’t do better/do so well on this particular test question on tornado development? Please explain.”</td>
<td>Please describe why you think you received this particular grade on this math test.</td>
<td>Open-ended</td>
<td></td>
</tr>
<tr>
<td>Self-Satisfaction</td>
<td>The level of satisfaction to the outcome on the math test.</td>
<td>Yes: “How satisfied are you with your score on this particular test item?”</td>
<td>How satisfied are you with your score on this math test?</td>
<td>1-7 Likert scale</td>
<td>1 = Not very satisfied 7 = Very satisfied</td>
</tr>
</tbody>
</table>

*Information regarding the group’s overall achievement was provided to students prior to the administering the self-reflection phase measures.
Phase Two: Stereotype Threat

Phase Two of the study was guided by the methods utilized in Good et al. (2008). In this phase, the final SRAT:MMA scale was administered to students in high and low stereotype threat conditions (See Appendix A for the IRB approval letter). Students were randomly assigned to the different stereotype threat conditions (high stereotype threat versus low stereotype threat). A 2 X 2 (i.e., high and low stereotype threat versus high and low self-regulation) quasi-experimental research design was employed with one between-subjects factor (stereotype threat) and one within-participants factor (self-regulation). The primary dependent variable was performance on the math test. Stereotype threat conditions were manipulated through test instructions and self-regulation was measured as an event with the SRAT:MMA. Essentially, this design allowed me to: 1) use the data investigate the initial validity of the modified microanalytic survey and 2) determine whether self-regulation processes as measured by the SRAT:MMA can alleviate stereotype threat in women.

Participants

A total of 191 undergraduate students enrolled a non-introductory math course participated in Phase Two of the study. Students were enrolled in one of the following classes: Math 114 (Analytic Geometry and Calculus II; N = 96 [50%]); Math 213 (Analytic Geometry and Calculus III; (N = 63 [33%]); Math 351 (Probability; (N = 22 [12%]); and Math 351 (Statistics; N = 10 [5%]). Non-introductory math courses were selected because students enrolled upper level courses are more likely to value math more than students enrolled in introductory courses (Good et al., 2008). Out of the total sample,
58.1% (N = 111) were males and 41.9% (N = 80) were female. The ethnic breakdown of the participants was as follows: 6.8% (N = 13) African American; 22.0% (N = 42) Asian; 41.9% (N = 80) Caucasian; 8.9% (N = 17) Hispanic; 7.9 % (N = 15) Middle Eastern; 4.7% (N = 9) Indian; 5.8% (N = 11) Mixed; 1.0% (N = 2) other; and 1.0% (N = 2) did not report their ethnicity. The average age of the participants was 20.74, with the oldest participant being 45, and the youngest participant being 18. In terms of college level, 32.5% (N = 62) of the participants were freshmen students, 30.4% (N = 58) were sophomore, 20.4% (N = 39) were juniors, and 15.2% (N =29) were seniors. Three students did not report their college level. In terms of the number of college math courses completed, 18% (N = 35) had completed at least one college math course, 40% (N = 78) had taken 2-3 courses, 21% (N = 38), had taken 4-5 courses, and 16% (N = 30) had taken 6 or more courses. Ten students did not report the number of math courses completed.

The target sample in Phase II of the study was women under stereotype conditions. Specifically, the goal of Phase II of the study was to investigate whether self-regulation processes could alleviate the stereotype threat effect on women’s achievement in math. Therefore, specific sample statistics regarding women are reported here. In this sample of 80 females, 50% (N = 40) of the sample were enrolled in Math 114, 23.8% (N = 19) were enrolled in Math 213, 18.8% (N = 15) were enrolled in Math 351, and 7.5% (N = 6) were enrolled in Math 352. The ethnic breakdown was as follows: 8.8% (N = 7) African American, 25% (N = 20) Asian, 43.8% (N = 35) Caucasian, 3.8% (N = 3) Hispanic, 5% (N = 4) Middle Eastern, 6.3% (N = 5) Indian, and 6.3% (N = 5) mixed or other. One student did not report her ethnic background. The average age of female
participants was 20.78, ranging from 18 to 45 years of age. In terms of college level, 32.5% (N = 26) were freshmen, 21.3% (N = 17) were sophomore, 22.5% (N = 18) were juniors, and 21.3% (N = 17) were seniors. Two students did not report their class level. Finally, most of the students had successfully completed at least three prior math courses at the college level (58.98% [N = 46]), and 41.02% (N = 32) had successfully completed four or more prior college math courses.

**Math Courses**

One of the assumptions of self-regulated learning is that the reason why students underperform is because they lack the self-regulation and motivation skills necessary to achieve (Cleary & Zimmerman, 2004). Through this perspective, lower level courses designed to prepare academically at-risk students for future courses would be the most ideal context to detect self-regulation deficiencies. In contrast, stereotype threat research suggests that the more students value doing well in a certain domain, the more pronounced the stereotype threat effect will be (Good et al., 2008; Spencer et al., 1999). This suggests that stereotype research should be conducted in upper-level math courses, where students are generally those who value and place more importance in achieving high academically than students enrolled in lower-level math courses.

The implications from these two lines of research (self-regulation and stereotype threat) have opposing, but theoretically- and empirically-based arguments for their conclusions. In order to bridge these two perspectives, upper and non-introductory level math courses were selected. There were two main rationales for focusing on an upper and non-introductory level math courses. First, microanalytic assessment measures are
particularly useful in diagnosing self-regulation deficiencies for at-risk student populations (Cleary et al., 2008). However, there is evidence that students enrolled in high level courses also experience deficiencies in self-regulated learning (Lynch, 2008). Therefore, this study will be able to conclude whether the SRAT:MMA can detect self-regulation deficiencies in upper level math students. Second, because stereotype threat is an important independent variable in Phase Two of the study, selecting upper level math courses may increase the chances of successfully manipulating the high and low stereotype threat conditions (Good et al. 2008).

**Selection Procedures**

A course search on Patriotweb showed that there were a total of eight upper level (course numbers within the 300 range) math courses offered in the Spring 2013 semester. All upper level instructors were contacted and recruited to participate in the study. Out of all of the eight professors contacted, only two professors agreed to participate on the condition that the experiment is conducted outside of official instruction time, or during reading week. To increase the sample size, professors who taught lower level math classes with prerequisites were also contacted. The classes selected were: Math 114 (Analytic Geometry and Calculus II) and Math 213 (Analytic Geometry and Calculus III). These classes were selected based on guidance from a math professor, suggesting that the last two calculus courses in the three course calculus sequence are more likely to have students who value math than other low level courses. Therefore, all of the Calculus II and III professors were contacted and three agreed to participate (two Calculus II
classes and one Calculus II class). This resulted in an additional 159 participants from the 32 participants from upper level math courses only.

To increase control for course differences, randomization occurred at the student level (which can limit course difference effects and potential confounding variables). Students enrolled in multiple math courses were controlled for by keeping their data from the first round of experimentation from their first class and discarding their data from the second round of experimentation. The results showed that not one student was enrolled in more than one of the courses selected for experimentation. Overall, four professors across five math classes agreed to participate (one professor, Teacher C, taught both Math 352 and Math 213). The final sample size breakdown by mathematics course and teacher was as follows: Teacher A Math 114: N = 31; Teacher B Math 114: N = 65; Teacher C Math 213: N = 63; Teacher C Math 352: N = 10; Teacher D Math 352: N = 22.

Task and Materials

**Self-regulation assessment tool.** This section will describe in detail the final Self-Regulation Assessment Tool: A Modified Microanalytical Approach (SRAT:MMA) administered to students during Phase Two of the study. The SRAT:MMA measured various processes under Zimmerman’s three phase model. These methods of measurement have shown strong levels of both reliability and predictive validity in measures assessing athletic (Kitsantas & Zimmerman, 2002) and academic tasks (DiBenedetto & Zimmerman, 2010, 2013), even with single item measures. Self-regulation was assessed as a task-specific event, where students test taking self-regulation was assessed before, during, and after the math test. See Appendix B for the full measure.
**Forethought phase measures.** Processes investigated in the forethought phase included: goal setting, strategic planning, self-efficacy, and task value. Goal setting and strategic planning fall under the task analysis category of forethought phase processes whereas self-efficacy and task value processes fall under the motivational beliefs category.

*Goal setting.* The goal setting items were developed following Zimmerman’s (2008b) description of advantageous properties of goals. First, students were asked to report the grade that they want to achieve on the math test they were about to take (out of five questions worth five possible points). Due to the specificity of this study (i.e., goal for a specific math test) only certain goal properties from Zimmerman (2008b) were adapted to fit the needs of this study. Out of the different goal properties that Zimmerman (2008b) list as advantageous, this study adopted: 1) goal difficulty (“On a scale of 1-7, how difficult will it be for you to achieve this grade on this test?”) and 2) process or outcome oriented goal (“Is it more important for you to focus on a) correctly following the steps to solving the problem or b) getting the answer right”). The responses for the goal challenge scale were measured on Likert scale ranging from 1 “Not very difficult” to 7 “Very difficult” whereas for the process/outcome goal measure, responses were coded as “1” = process goal, “0” = outcome goal.

*Self-efficacy.* The self-efficacy item was adapted from DiBenedetto and Zimmerman (2010), which was, “On a scale of 1 – 7, how confident are you that you will be able to reach the grade you want to achieve on this math test?” anchored on 1 (“Not very confident”) to 7 (“Very Confident”) Likert scale.
Strategic planning. Strategic planning was assessed with the following open-ended question: “What strategies are you planning on using to help you succeed on this test? For example, any test taking strategies, techniques, methods, or line of attacks? Please describe them below.” This item was adapted from DiBenedetto and Zimmerman (2010). Student responses were scored by counting the number of strategies listed by two trained researchers. For example, one student described: “Process of eliminating multiple choice answers that are clearly incorrect, right away.” This was coded as having one strategic plan. Another student listed: “Answer ones I immediately know how to do. Skipping those I do not, then going back to answer. Also take a break in the middle by putting pencil down and gathering myself,” which was coded as two strategies. Any disagreement was discussed until inter-rater reliability reached 100%.

Task value. Test-specific task value was assessed with the following question adapted from the Self-and Task Perception Questionnaire (STPQ: Eccles & Wigfield, 1995): “How important is it for you to do well on this math test?” Student responses were measured on a Likert scale ranging from 1 “Not very important” to 7 “Very important”. Additionally, domain-specific task value was assessed with the following question: “How important is it for you to be good at math?” which was also measured on a Likert scale ranging from 1 “Not very important” to 7 “Very important”. The anchors on the scale was modified from the original STPQ from “Not at all important” to “Not very important” to make the different anchors on this specific measure consistent with each other.
**Performance phase measures.** Processes investigated in the performance phase included attention focusing strategies as part of the self-control category of processes and metacognitive monitoring as part of the self-observation processes. Questions under this phase were administered while students were completing the math test, except for attention focusing strategies, which were assessed immediately after students have completed the math test.

**Metacognitive monitoring.** Questions assessing students’ metacognitive monitoring were administered before and after each test question. The question prior to the math item was “*How confident are you about your understanding of this question?*” (α = .77) and the question after the math item was “*How confident are you that you have answered this question correctly and completely?*” (α = .66). The items were adapted from DiBenedetto and Zimmerman (2010) and were measured on a 1 (Not very confident at all) to 7 (Very confident) Likert scale.

The confidence judgements that are typically used to measure metacognitive monitoring in students (Moores et al., 2006; Nietfeld et al., 2005) overlaps considerably with self-efficacy (i.e., confidence in one’s ability to achieve a certain goal; Bandura, 1997). Although these two constructs do covary, there are unique differences (Moores et al., 2006). For example, self-efficacy is a broader construct that encompasses students’ overall feelings of whether they can achieve a goal, regardless of how accurate they calibrate their actual abilities. Metacognitive monitoring, on the other hand, includes a judgement component (Kelemen, 2000) of whether or not learning has occurred, or even as a feeling like they know the answer (Metcalfe et al., 1993). Therefore, self-efficacy is
measured as a broader construct and metacognitive monitoring is measured as an item-specific “confidence in answer” judgement in this study.

Attention Focusing Strategies. Two questions (one closed- and one open-ended) were administered to measure students’ attention focusing strategies: “At any point during the test, did your concentration wander?” (yes/no response); and “If yes, what did you do to refocus your concentration?” All of the student responses to this measure were scored by two trained researchers by counting the number of strategies listed. For example, one student responded: “Tried to remember the importance of this test” which was coded as one strategy, another student responded “I calmed down and reread the question” which was coded as two strategies. Two trained raters determined the number of strategies listed for each student and any disagreement was discussed until inter-rater reliability reached 100%.

Self-reflection phase measures. Processes investigated under this phase fell under two categories: self-judgment and self-reaction. Self-evaluation and causal attributions were assessed as self-judgment processes whereas self-satisfaction was assessed as a self-reaction process. These measures were administered after students completed the math test and received the final test grade.

Self-evaluation. Students were asked to self-evaluate their learning based personal mastery criteria, prior performance, or social comparison. To provide students with the option to self-evaluate using social comparison standards, the breakdown of the class test performance was written on the board after all students had completed the test. Self-evaluation was then assessed with the following closed ended item: “Please complete the
following sentence: “I feel that my performance on this test was successful/unsuccessful because (Circle a, b, or c).” Students were able to select either a) “I did/didn’t do better than the class average”; b) I did/didn’t do better than on my previous tests; or c) “I did/didn’t meet my goal.” This question was designed to assess whether the students evaluated their learning by personal mastery criteria, past performance, or against the performance of their peers. Upon collecting the data, the responses were coded as “1” = social comparison, “2” = prior performance, and “3” = personal mastery criteria.

The coding scheme for self-evaluation was designed to align with prior research suggesting that the most adaptive form of self-evaluation is based on personal mastery criteria, followed by prior performance and social comparison (Schunk & Ertmer, 1999; Zimmerman, 2008a). Therefore, although this variable was technically a three level categorical variable, the coding scheme mirrored a three level continuous variable anchored on a “3” most effective form of self-evaluation to “1” least effective form of self-evaluation. The coding scheme of this variable was maintained when it was entered into the two composite self-regulation variables: self-reflection phase composite and overall self-regulation composite.

Causal attributions. The causal attributions measure regarding student performance was adapted from DiBenedetto and Zimmerman (2010). Specifically, students were asked one question after they received their test grade: “Please describe why you think you received this particular grade on this math test.” Student responses were coded based on the levels of attributions (i.e., stability (1 = stable, 0 = unstable), controllability (1 = controllable, 0 = uncontrollable), and locus of control (1 = internal, 0
Weiner (1986) suggested that overall, attributions to achievement tend to fall along the lines of ability, effort, task difficulty, and luck. Ability would be a stable, uncontrollable, and internal attribution; effort would be an unstable, controllable, internal attribution; task difficulty would be an unstable, uncontrollable, external attribution; and finally luck would be an unstable, uncontrollable, and external attribution. All of the student attributions were coded based on this framework. Two trained raters coded the all of the student attributions according to the framework and discussed any disagreement until there was 100% consensus.

It is important to discuss here that the stability dimension of attributions were originally coded by “1” = unstable and “0” = stable. This was done because Weiner (1986) suggested that unstable attributions or attributions that change over time are more associated with higher levels of academic performance than attributions that do not change over time. However, the stability dimension of attributions is closely related to expectancies for success (Schunk et al., 2008) and does not uniformly negatively affect all students (Gibb et al., 2002). Specifically Gibb et al. (2002) found that internal and stable attributions tend to result in higher levels of achievement when students typically have had prior experiences of success within a given domain. This was the case in this specific study, where students must have had prior experiences of success in math given that they were enrolled in a non-introductory course that required them to have successfully completed at least one prior math course. Therefore, the stability dimension of attributions were coded as “1” = stable and “0” = unstable. The locus of control (1 = internal, 0 = external) and controllability (1 = controllable, 0 = uncontrollable)
dimensions of attributions were maintained, given that prior research showed that internal and controllable attributions result in higher levels of achievement than external and uncontrollable attributions, regardless of prior performance (Gibb et al., 2002) Self-satisfaction. The self-satisfaction measure was adapted from DiBenedetto and Zimmerman (2010). After completing the test and receiving the score, students were asked “How satisfied are you with your score on this math test?” Students responded on a Likert scale ranging from 1 (“Not very satisfied”) to 7 (“Very satisfied”).

Achievement. Achievement was measured with a professor created math test. The math test was unique to each class and was referred to as a practice test for an upcoming final exam. This achievement measure was used as: 1) an initial indicator of validity of effective versus ineffective self-regulation practices (e.g., does high self-efficacy beliefs as measured by the SRAT:MMA actually result in higher levels of achievement than low self-efficacy beliefs?), 2) a moderator of the relationship between different phases of self-regulation, 3) the dependent variable in the predictiveness of the self-regulation processes as measured by the SRAT:MMA, and 4) an indicator of stereotype threat performance differences. Although it would have been ideal to collect multiple indicators of achievement (e.g., performance on prior math tests, overall course grades, and performance on the upcoming final), this was not possible given the conditions of anonymity.

In prior math stereotype threat studies, achievement measures were mostly created by researchers that consisted of GRE or SAT math questions. However, to maintain the naturalistic context, the professors were asked to develop a math test that
they would typically administer to their students. All professors were asked to develop a math test as an optional extra credit practice test for the upcoming math final. Professors were given the option of creating a short answer or a multiple choice math test. However, professors were requested to create a test that: 1) would be easily graded by individuals without a strong math background; 2) would take no more than 25-30 minutes for students to complete; 3) consist of five questions; and 4) would be similar to the types of tests typically given in class. Overall, one math test was short answer whereas all the other four tests all had five multiple choice questions. See Appendix C for all of the tests. Student performance on this math test was the primary outcome and achievement measure. The math test was administered after the stereotype threat manipulation and in conjunction with the modified microanalytic assessment of self-regulated learning.

**Test control procedures.** Three procedures were followed to ensure that the professor created math tests were relatively similar in terms of difficulty level. First, all professors were asked to create a five question 30 minute test. Second, all tests were reviewed by two math experts to ensure difficulty levels were comparable. Third, randomization of the stereotype threat conditions occurred at the student level to ensure that test and course differences were evenly spread across high/low stereotype conditions. An additional question was included at the end of the survey as a test control strategy: “How difficult was this test” which was measured on a 1 (“Not very difficult) to 7 (“Very difficult”) Likert scale. Ultimately, perceived test difficulty levels did not differ as a function of math test, $F(4, 186) = 1.92, p = .11, \eta^2_p = .04$. 


Demographic questionnaire. This questionnaire collected student demographic information which included age, gender, major, number of math classes completed, and ethnic background.

Manipulation check. To confirm the validity of the stereotype threat manipulations, students were asked to respond to the item, “The test was biased” on a 1 “not at all” to 15 “extremely” Likert scale (Good et al., 2008).

Administration Procedures

Experimentation began during University reading week (~week of May 5th 2013), or one week before students’ scheduled final exam. One week prior to the experiment, the professor announced to their class that there was an extra credit opportunity during reading week. Students who chose to meet during the scheduled day, time, and place would be eligible for a certain amount of extra credit that would count towards their final exam or grade. Students were told that the amount of extra credit that they receive will be dependent on how well they do on the test; however, in reality all students who completed the test received the same amount of extra credit.

Although there is evidence that students tend to exert more effort and self-regulation for consequential tests (Sundre & Kitsantas, 2004) than for non-consequential tests, the test administered counted as extra credit as opposed to the actual grade because of ethical issues. Specifically, because students were placed under stereotype threat conditions, their performance may not be a reflection of their true abilities, especially for women. Therefore, it would have been unethical count the actual performance on this math test towards students’ final grade. The following sections will first describe the
Randomization procedures, followed by a description of the high and low stereotype conditions. Finally, the last section will describe the survey administration procedures.

**Randomization procedures.** Students were randomly placed into high and low stereotype conditions based on gender to ensure that an evenly distributed number of males and females were spread across conditions. For example, if there were ten men and six women in the class, I would have ensured that there were five high and five low stereotype threat conditions for the men, and three high and three low stereotype conditions for the women. In order to do so, the survey packets were mixed together in an “AB” pattern (one high stereotype threat packet, one low stereotype threat packet). When the packets were handed out, males were handed a packet from the top of the pile and females were handed a packet from the bottom of the pile.

**High stereotype threat condition.** Stereotype threat conditions followed the same procedures as outlined by Good et al. (2008). Students were administered a math test described by the instructor as a practice test for an upcoming course exam. The math problems included material that was similar to that on the final exam. Prior to administering the test, the professors of the course explained to the students that the practice test will help them prepare for the upcoming examination and encouraged them to take the test seriously. To invoke stereotype threat, students read the following statement taken verbatim from Good et al. (2008), which has been shown to successfully invoke stereotype threat:

> For the next X minutes, you will be taking a math test aimed at measuring your mathematical abilities. Why? As you probably know, math skills are crucial to
performance in many important subjects in college. Yet surprisingly little is known about the mental processes underlying math ability. This research is aimed at better understanding what makes some people better at math than others. After you finish the test, we will score it. This will enable us to analyze your performance and compare it with other students taking this test. (Good et al., 2008, p. 21)

**Low stereotype threat condition.** Procedures for the low stereotype threat condition followed the same procedures as the students in the high stereotype threat condition. However, the test instructions further included the following statement taken verbatim from Good et al. (2008), which has been shown to eliminate stereotype threat:

> What about gender differences? This math test has not shown any gender differences in performance or math ability. The test has been piloted in many math courses across the nation to determine how reliable and valid the test is for measuring math ability. Analysis of thousands of students' test results has shown that males and females perform equally well on this test. In other words, this math test shows no gender differences. (p. 22)

This approach to manipulating stereotype threat from Good et al. (2008) differ from traditional manipulation methods. Specifically, popular methods of manipulating stereotype threat in previous studies (e.g., Quinn & Spencer, 2001; Spencer et al., 1999) were to either implicitly or explicitly suggest that the math test was gender biased or raise the notion that females generally do worse in math as compared to males. However, these methods of stereotype threat manipulation are not representative of the classroom setting.
Students who take diagnostic math tests are typically not told that the test might be gender biased. Therefore, an alternative approach to manipulating stereotype threat is to simply suggest that the math test is diagnostic of math ability, with no reference to gender differences or gender related issues.

Another important distinction in this study and other stereotype threat studies was the way stereotype threat was nullified. In many previous stereotype threat studies, researchers removed stereotype threat by removing the diagnosticity of the test (Good et al., 2008). For example, in low stereotype threat conditions, participants would be told that the math test would not be diagnostic of their mathematical abilities or that the test only measured a general cognitive process, like problem solving. However, this type of manipulation does not align with actual testing situations in the college classroom. That is, students are well aware that course examinations are meant to diagnose their current abilities and that their performance on those exams contributes to their final achievement scores. Therefore, to accurately align stereotype manipulations to the real world setting, the diagnostic nature of the math test was maintained for the low stereotype threat condition.

**Survey and testing procedures.** Survey packets were separated into seven sections: 1) informed consent (see Appendix D), 2) high/low stereotype threat manipulation, 3) forethought phase measures, 4) performance phase measures along with the math test, 5) self-reflection phase measures, 6) demographic questionnaire, and 7) debriefing form (See Appendix E). The following section will describe how the survey was administered.
Three trained researchers assisted in administering surveys and collecting data. On the day of the experiment, the professor introduced me as the researcher. Afterwards, the following introduction was read to each class:

Hello, my name is Faye Huie and I am a doctoral student in the College of Education and Human Development. I am here today to collect data for my dissertation about the mental processes underlying mathematical abilities. Only students who are 18 years old and older are eligible to participate. If you are under 18 years of age, please let me know now. If you choose to participate, I will ask you to complete a series of surveys which will take approximately 5-10 minutes to complete. Additionally, you will be asked to take a math test which will help give you and your professor a better idea of your strengths and weaknesses for your upcoming final. You will be taking the math test today regardless of whether or not you choose to participate in this study and the amount of extra credit you receive will be dependent upon your performance on this math test.

My colleagues are now handing out the surveys and math tests. Please do not flip through the survey when you receive it.

After survey packets were all handed out, students were asked to review and sign the two copies of informed consents with the following directions read by me:

Here in the first two pages, you will find the informed consent. This form tells you about the research and what your participation will entail. Just to highlight a few things, your participation is completely voluntary and your identity will be anonymous. I will not collect any identifying information but you will be
assigned an ID number. That number is highlighted at the top of the informed consent as well as your math test that you will be taking later. If you choose to participate, please sign both copies of the informed consent and rip off the top one for your own records. Make sure you keep the informed consent because you will need it later to help you identify which test is yours to pick up. Any questions?

There were no questions from students in any of the classes. After everyone had finished reading and signing the informed consents, the following directions were read to the students:

So now, I’m going to read out some instructions before you complete the survey and test. Please listen carefully. First, everyone should read page 2 of the packet which has a little information about your test today. Afterwards, complete the next page, which is about your motivation in math. After you complete that part of the survey, you will then take the actual math test. There are five test questions. You will be asked one question before and after you answer each math question. Please answer those questions on a 1-7 scale and circle the number that best represents how you feel. Please also use the separate sheet of paper that is provided on page 6 for any written work. You may not use calculators. When you are done with your test, please flip around the packet and raise your hand. One of us will come and grade your test. You may now complete the first part of the survey and math test. You have X minutes and your time starts now.
After these directions were read aloud, students then read the high and low stereotype manipulation statement. To ensure that students followed the directions, the directions on the bottom of each page were bolded, telling students to continue to the next page until it came to the end of the math test. At the end of the math test, there was a bolded statement that read: “STOP HERE: Please flip over the entire survey packet and raise your hand once you are done with the test. One of us will come and grade it. Do not continue with the survey until further instructions.”

The research assistants and I walked around the room while students were testing and began to grade the tests as they were completed. As we were grading, we also recorded the grades. Specifically, we tabulated the frequency of each score. Once all tests had been graded and time was up, the class score breakdown was tallied and written on a board in in front of the class. This was done so that the students could answer the self-evaluation question, regarding how they judge the success or failure of their performance by social comparison, prior performance, or personal goals. After the breakdown was written on the board, the following instructions were read: “So on the board I have a breakdown of how the class scored on the test. Go ahead and complete the rest of the survey to the end.”

After all students had completed the self-reflection phase survey, demographic questionnaire, and read the debriefing form, the following closing statement was read:

Thank you everyone for your participation in this research. As you all have read, this study was really about stereotypes and how it influences women’s math test performance. Please email me to let me know if you have any questions and I will
be happy to answer them. We will now come and collect the survey packets. I am going to go make copies of your completed and graded tests. You may pick up your test in your professor’s mailbox located in Room X in X OR, I will bring back your tests after I have photocopied them. Use your ID number to identify which test is yours. Thank you again.

Survey packets were then collected. Immediately after collecting the survey packets, the research assistants and I pulled out the tests and photocopied them. Depending on the professor’s directions for returning the tests back to students, the tests were either dropped off in the professor’s mailbox or brought back to the classroom.

**Manipulation Check Results**

The manipulation check was guided Good et al. (2008), where a 2 (stereotype threat manipulation) x 2 (gender) ANOVA with test bias perceptions as the dependent variable was performed to investigate the fidelity of the manipulation. Treatment fidelity is confirmed when a stereotype threat main effect is detected. Specifically, students in the high stereotype threat condition should report higher perceptions of bias than students in the low stereotype threat condition, regardless of gender.

A meta-analysis conducted by Nguyen and Ryan (2008) found that the overall effect size (Cohen’s $d$) in gender stereotype threat studies was .25. A power analysis conducted by the G Power software showed that the sample size required in order to detect a large effect size ($d = .50$) at an 80% power level was 48. The sample required for a medium effect size ($d = .30$) was 126, and the sample required for a small effect size ($d = .10$) was 1095. Based on the meta-analytic findings from Nguyen and Ryan (2008), the
sample size required to detect a Cohen’s $d$ of .25 at 80% power was 179. The sample size in this study was 191; therefore, the sample size provided sufficient power to detect a large or medium gender effect in stereotype threat. The ANOVA results revealed no statistically significant main stereotype threat effect, indicating that the stereotype manipulation was not successful ($F(1, 190) = .17$, $p = .68$, $\eta^2_p = .002$). Further, the critical test for stereotype threat, using student achievement on the math test as the dependent variable and gender (male and female) and stereotype threat (high and low stereotype threat) as the between-subjects factor showed no significant gender by stereotype threat interaction, $F(1, 191) = .53$, $p = .47$, $\eta^2_p = .003$, indicating that males and females performed at similar levels regardless of the stereotype threat condition.

Prior research suggests that women are more prone to stereotype effects when their ability is questioned in a domain they value (Good et al., 2008). In this study, it was expected that women who care about math and felt that the test was difficult would have been more likely to experience stereotype threat effects than women who do not care about their performance in math. Therefore, efforts were made to investigate whether the manipulation worked on a specific set of students who valued math and who felt that the test would be difficult.

The overall sample was split based on several demographic and self-regulatory variables related to math value and test difficulty. The demographic variables included class level (i.e., class level: upper level (juniors and seniors) students tend to take their performance more seriously than lower level students), math-related major (i.e., students who report having a math related major are more likely to value their performance in
math as compared to students who do not report having a math related major), and professor (i.e., qualitative evidence suggested that some teachers emphasized the importance of the test more so than others). The self-regulatory variables included math value, test value, test effort, test difficulty, and meaningful attributions to test performance (e.g., students who blatantly reported that they did not take the test seriously because they were not being graded on it versus students who reported that they need more time to study).

After splitting the sample based on the demographic and self-regulatory variables, the results were as follows:

Demographic variables:

a) Class level: (N = 32): Only students enrolled in upper level math courses (Math 351 and Math 352) were included in these analyses. The results showed no statistically significant stereotype threat effect, indicating that the manipulation did not work on students who were enrolled in upper level math courses ($F(1, 28) = .04, p = .84, \eta^2 = .001$).

b) Class level (including mid-level math class) (N = 95): These results were based on a sample of students who were enrolled in upper level math courses (Math 351 and Math 352) or the mid-level math course (Math 213). The results revealed no statistically significant stereotype threat effect, indicating that the
stereotype threat manipulation did not work on students enrolled in upper level math classes \( (F(1, 91) = .04, p = .84, \eta^2_p < .001) \).

c) Math related major (N = 166): These results were based on a sample of students who were had reported a math related major. Some of the examples of math related majors included physics, engineering, and accounting. The results revealed no statistically significant stereotype threat effect, indicating that the stereotype threat manipulation did not work on students with a math related major \( (F(1, 162) = .001, p = .86, \eta^2_p < .001) \).

d) Completed three or more math courses (N = 104): Students who reported completing at least three or more math classes were selected to be included in these analyses. The results revealed no statistically significant stereotype threat effect, indicating that the stereotype manipulation was not successful on students who had completed at least three math courses \( (F(1, 100) = .05, p = .83, \eta^2_p < .001) \).

e) Teacher: Individual ANOVAs were run for each of the four participating professors. The results indicated that the stereotype threat manipulation was not successful in any of the professors’ classes.

1. Professor A: N = 31; \( F(1, 27) = .43, p = .52, \eta^2_p = .02 \).
2. Professor B: N = 65; \( F(1, 61) = .04, p = .83, \eta^2_p = .001 \).
3. Professor C: N = 22; \( F(1, 18) = .16, p = .69, \eta_p^2 = .01. \)

4. Professor D: N = 73; \( F(1, 69) = .02, p = .88, \eta_p^2 < .001. \)

**Self-regulatory variables.** Students who reported high perceptions of specific self-regulation variables related to math value were also specifically selected to investigate whether the manipulation worked on this subset of students. For example, students who made a stronger effort to do well on the test were expected to value math more so than students who did not make a strong effort to do well on the test. Therefore, high levels of the following variables were specifically selected (i.e., one standard deviation above the mean): math value, test value, test effort, and test difficulty. Individual stereotype threat by gender ANOVAs were then performed for each of the selected variables. The results were as follows:

a) Math value (N = 65): \( F(1, 61) = .01, p = .94, \eta_p^2 < .001. \)

b) Test value (N = 76): \( F(1, 72) = .18, p = .67, \eta_p^2 = .002. \)

c) Test effort (N = 36): \( F(1, 32) = .30, p = .59, \eta_p^2 = .01. \)

d) Test difficulty (N = 42): \( F(1, 38) = .45, p = .51, \eta_p^2 < .001. \)

e) Meaningful attributions (N = 173): \( F(1, 169) = 1.02, p = .32, \eta_p^2 = .01 \)

There was also evidence in research literature that women who are moderately identified in mathematics are the most susceptible to stereotype threat (Nguyen & Ryan, 2008). Specifically, in a meta-analysis, Nguyen and Ryan (2008) found that the stereotype threat effect size was stronger for moderately math identified women than for highly identified women, suggesting that there may be a stereotype reactance (i.e., individuals faced with a
blatant negative stereotype may react strongly against it because they perceive it as a source of oppression, thereby invoking behaviors to overtly go against the stereotype) effect among highly math identified women. Therefore, the middle third of the responses to math value (one standard deviation above and below the mean excluded) were selected for fidelity analyses \((N = 105; F(1, 101) = .02, p = .90, \eta_p^2 < .001)\), which also failed to confirm the fidelity of the manipulation.

Ultimately, two factors, level of math class and lack of consequences on the math test may have contributed to the inability to confirm the stereotype threat manipulation. Another contributing factor may have been the way in which stereotype threat was manipulated, where the low stereotype threat manipulation deviated from traditional methods. Specifically, traditional methods of establishing a low stereotype threat condition typically involve eliminating the diagnosticity of the test. However, because this study attempted to manipulate stereotype threat under real-world circumstances, the diagnosticity of the test was maintained and a statement indicating the absence of gender differences was added (Good et al., 2008). Although Good et al. (2008) were able to confirm the manipulation, perhaps it was the combination of the level of math class and lack of test consequences that contributed to the inability to manipulate stereotype threat in this study. Another possibility may be that students simply did not read the stereotype threat manipulation. Because the fidelity was not confirmed, this study will focus primarily on the primary research question, which was: Can the SRAT:MMA diagnose college student self-regulated learning at the classroom level during test taking in a math course?
Chapter Four

The purpose of this study was to investigate the initial validity of the SRAT:MMA. The secondary purpose was to investigate whether or not the processes as measured by the SRAT:MMA can detect self-regulation deficiencies under stereotype threat conditions. As a result, the following research questions were formulated:

1. Can the SRAT:MMA diagnose college student self-regulated learning at the classroom level during test taking in a math course?
   a. Do students with high self-regulation perform better than students with low self-regulation on a math test?
   b. Does the relationship between the different phases of self-regulation vary across levels of student achievement on a math test?
   c. Can the self-regulation variables, as measured by the SRAT:MMA predict college student achievement on a math test?

2. Can self-regulation processes, as measured by the SRA-MMA, alleviate stereotype threat effects in women enrolled in undergraduate math courses?

The following sections will first describe how composite measures were formed. Following, the descriptive analyses (i.e., overall descriptive statistics and intercorrelations) will be reported. Finally, the last sections will report each of the results according to the research questions and sub-questions.
Formation of Composite Measures of Self-Regulation

Four composite measures of self-regulation were formed: forethought phase, performance phase, self-reflection phase, and overall self-regulation. First, all individual self-regulation processes were standardized by transforming them into z-scores. The standardized variables were created only for the purposes of generating composite measures of self-regulation. Unstandardized scores were used with all analyses investigating individual self-regulation processes. The composite measures were formed by computing the mean of the standard scores for each student. Similar approaches to creating a composite scale of self-regulation have also undertaken by Kitsantas (2002), Kitsantas and Zimmerman, (2002) and Zimmerman and Kitsantas (2014). Grade goal, goal difficulty, process versus outcome goals, self-efficacy, strategic planning strategies, test-specific task value, and math domain task value were combined to create the forethought phase composite. Metacognition: understanding of question, metacognition: confidence in answer, and attention focusing strategies were combined to create the performance phase composite. Self-evaluation, attributions: stability, attributions: controllability, attributions: locus of control (internal vs. external), and self-satisfaction were combined to create the self-reflection composite. Finally, all of the individual variables mentioned above were combined to create the overall self-regulation composite.

In terms of the self-evaluation variable, although it was technically a three level categorical variable (i.e., mastery criteria, prior performance, or social comparison), the coding of this variable aligned with prior research which suggests that the most effective self-evaluation criteria is personal mastery criteria, followed by prior performance and
social comparison (Schunk & Ertmer, 1999; Zimmerman, 2008a). As a result, the coding scheme imitated a three level continuous variable anchored on a “3” (most effective form of self-evaluation) to “1” (least effective form of self-evaluation) Likert scale. This variable was then entered as a continuous variable in two composite measures of self-regulation (self-reflection phase and overall self-regulation) after it was standardized into z scores.

The attributions variable was separated into three variables before adding them into two composite measures of self-regulation: overall self-reflection phase processes and overall self-regulation. The three variables included stability (1 = stable, 0 = unstable), controllability (1 = controllable, 0 = uncontrollable), and locus of control (1 = internal, 0 = external). Although Weiner (1986) suggests that unstable attributions are more associated with performance than stable attributions, this relationship is not consistent between different groups of students (Gibb et al., 2002). Specifically, Gibb et al. (2002) found that stable attributions are associated with higher achievement when students have historically done well in a given domain than when students have not historically done well in a given domain.

In this study, all students (except for the ten students who did not report the number of college math classes they had successfully completed) have had prior experiences of success in math, suggesting that this sample of students may not be negatively influenced with stable attributional styles. Statistical analyses showed that this was indeed the case. More stable attributions were more associated with higher levels of achievement than less stable attributions (stable attributions: $M = 4.24$, $SD = 2.55$;
unstable attributions: $M = 2.55, SD = 1.30; t(167) = 5.06, p < .001, d = 1.30$). Therefore, to align the coding scheme with prior research, the stability dimension of attributions were coded as “1” = stable and “0” = unstable

**Descriptive Analyses**

See Tables 2 and 3 for the overall descriptive (unstandardized) statistics and intercorrelation tables. In terms of the descriptive statistics, the means and standard deviations for each individual self-regulatory process were computed for the overall sample and by high achievement (a score of three out of five possible points or higher) and low achievement (a score of two out of five possible points or lower). These high and low achievement markers were selected to broadly explore the potential general patterns of self-regulation with the entire sample. The binary variables were process versus outcome goals and the three dimensions of attributions (i.e., stability, controllability, and locus of control). The means of these variables should then be interpreted as the percent of those who were coded as “1”. Overall, students reported rather sufficient levels of self-regulation. For example, students overall reported high grade goals ($M = 4.60, SD = .70$), math domain task value ($M = 5.49, SD = 1.52$), and self-efficacy ($M = 4.21, SD = 1.40$). In terms of achievement levels, there were some numerical differences between high and low achieving students. For example, students who performed high reported slightly higher grade goals ($M = 4.64, SD = .65$) and slightly lower goal difficulty ($M = 4.00, SD = 1.31$) perceptions than students who performed low (grade goal: $M = 4.55, SD = .77$; goal difficulty: $M = 4.54, SD = 1.50$). Additionally, students who achieved high reported slightly higher self-efficacy perceptions ($M = 4.52, SD = 1.30$) and math domain task
value ($M = 5.79, SD = 1.39$) than students who achieved low (self-efficacy $M = 3.80, SD = 1.42$; math domain task value $M = 5.11, SD = 1.62$) Overall, these initial descriptive statistics provide some insight that although this sample reported relatively strong self-regulation; some achievement level differences may exist.

In terms of the intercorrelations, there were several interesting patterns. First, self-satisfaction seemed to contain the most statistically significant correlations with other self-regulation variables. Out of the 17 total variables in the intercorrelation table, self-satisfaction was significantly related to 11 of the variables. These variables included achievement ($r = .49, p < .01$), goal difficulty ($r = -.25, p < .01$), self-efficacy ($r = .26, p < .01$), math domain task value ($r = .14, p < .05$), metacognitive monitoring: understanding of question ($r = .31, p < .01$), metacognitive monitoring: confidence in answer ($r = .36, p < .01$), attentional control ($r = .14, p < .05$), self-evaluation: social comparison ($r = .19, p < .01$), self-evaluation personal goal ($r = -.15, p < .05$), attributions: stability ($r = .40, p < .01$), and attributions controllability ($r = -.22, p < .01$). Although most of these correlations were rather weak, it may suggest that self-satisfaction is an important component in self-regulation, particularly in that it may play a cyclical role in informing other self-regulation processes.

Other patterns also showed that goal difficulty was generally negatively associated with other self-regulation processes like self-efficacy ($r = -.55, p < .01$) and the two measures of metacognition (understanding of question: $[r = -.36, p < .01]$ and confidence in answer $[r = -.37, p < .01]$). Self-efficacy, on the other hand, was generally positively associated with other aspects of self-regulation such as math test task value ($r$
= .17, \( p < .05 \)), and math domain task value \( (r = .30, p < .01) \). Overall, these descriptive analyses provide some preliminary insight on the patterns of self-regulation within this sample of students.
Table 2

*Overall Self-Regulation and Self-Regulation by Achievement Level Means and Standard Deviations*

<table>
<thead>
<tr>
<th>Self-Regulation Variables (Scale: Minimum and Maximum Score)</th>
<th>Overall Sample N = 191</th>
<th>High Achievement N = 108</th>
<th>Low Achievement N = 83</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>Achievement (0-5)</td>
<td>2.74</td>
<td>1.39</td>
<td>3.79</td>
</tr>
<tr>
<td>Grade Goal (2-5)</td>
<td>4.60</td>
<td>.70</td>
<td>4.64</td>
</tr>
<tr>
<td>Goal Difficulty (1-7)</td>
<td>4.24</td>
<td>1.42</td>
<td>4.00</td>
</tr>
<tr>
<td>Process vs Outcome Goal (1 = Process goal, 0 = Outcome goal)</td>
<td>.71</td>
<td>.46</td>
<td>.71</td>
</tr>
<tr>
<td>Self-Efficacy (1-7)</td>
<td>4.21</td>
<td>1.40</td>
<td>4.52</td>
</tr>
<tr>
<td>Strategic Planning (0-4)</td>
<td>.63</td>
<td>.92</td>
<td>.60</td>
</tr>
<tr>
<td>Task Value: Math Test (1-7)</td>
<td>4.79</td>
<td>1.88</td>
<td>4.76</td>
</tr>
<tr>
<td>Task Value: Math Domain (1-7)</td>
<td>5.49</td>
<td>1.52</td>
<td>5.79</td>
</tr>
<tr>
<td>Metacognition: Understanding of Question (1-7)</td>
<td>4.55</td>
<td>1.32</td>
<td>4.81</td>
</tr>
<tr>
<td>Metacognition: Confidence in Answer (1-7)</td>
<td>4.08</td>
<td>1.30</td>
<td>4.46</td>
</tr>
<tr>
<td>Attentional Control (0-3)</td>
<td>.39</td>
<td>.60</td>
<td>.45</td>
</tr>
<tr>
<td>Self-Evaluation (1 = Social comparison; 2 = Prior performance; 3 = Mastery criteria)</td>
<td>2.48</td>
<td>.79</td>
<td>2.46</td>
</tr>
<tr>
<td>Attributions: Stability (0 = Unstable; 1 = Stable)</td>
<td>.10</td>
<td>.30</td>
<td>.16</td>
</tr>
<tr>
<td>Attributions: Controllability (0 = Uncontrollable; 1 = Controllable)</td>
<td>.80</td>
<td>.40</td>
<td>.79</td>
</tr>
<tr>
<td>Attributions: Internal vs External (0 = External; 1 = Internal)</td>
<td>.86</td>
<td>.35</td>
<td>.85</td>
</tr>
<tr>
<td>Self-Satisfaction (1-7)</td>
<td>3.38</td>
<td>2.14</td>
<td>4.49</td>
</tr>
</tbody>
</table>

*Note.* The means of binary variables should be interpreted as the percentage of the sample who were coded as “1.”
### Table 3

**Intercorrelations between all Individual Self-Regulation Variables**

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Achievement</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Grade Goal</td>
<td>.08</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Goal Difficulty</td>
<td>-.25**</td>
<td>.08</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Process vs Outcome Goal</td>
<td>.08</td>
<td>-.07</td>
<td>-.02</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Self-Efficacy</td>
<td>.33**</td>
<td>.14+</td>
<td>-.55**</td>
<td>.06</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Strategic Planning</td>
<td>-.05</td>
<td>.04</td>
<td>.02</td>
<td>.02</td>
<td>-.01</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Task Value: Math Test</td>
<td>-.01</td>
<td>.23**</td>
<td>.06</td>
<td>.13+</td>
<td>.17*</td>
<td>.18*</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Task Value: Math Domain</td>
<td>.22**</td>
<td>.16+</td>
<td>-.12+</td>
<td>.08</td>
<td>.30**</td>
<td>.02</td>
<td>.31**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Metacognition:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Understanding of Question</td>
<td>.27**</td>
<td>.28**</td>
<td>-.36**</td>
<td>.08</td>
<td>.43**</td>
<td>.02</td>
<td>.11</td>
<td>.28**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Metacognition:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Confidence in Answer</td>
<td>.38**</td>
<td>.27**</td>
<td>-.37**</td>
<td>-.03</td>
<td>.45**</td>
<td>.02</td>
<td>.10</td>
<td>.25**</td>
<td>.75**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Attentional Control</td>
<td>.10</td>
<td>.07</td>
<td>.06</td>
<td>-.11</td>
<td>-.07</td>
<td>-.09</td>
<td>-.06</td>
<td>-.04</td>
<td>-.07</td>
<td>-.06</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. Self-Evaluation:</td>
<td>.04</td>
<td>.03</td>
<td>.02</td>
<td>-.12</td>
<td>-.11</td>
<td>-.11</td>
<td>-.01</td>
<td>-.08</td>
<td>.05</td>
<td>.01</td>
<td>.03</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social Comparison</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. Self-Evaluation:</td>
<td>.05</td>
<td>-.02</td>
<td>.01</td>
<td>.04</td>
<td>.12+</td>
<td>.10</td>
<td>.14+</td>
<td>.10</td>
<td>.03</td>
<td>.03</td>
<td>-.03</td>
<td>-.19**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prior Performance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. Self-Evaluation:</td>
<td>-.07</td>
<td>-.01</td>
<td>-.02</td>
<td>.07</td>
<td>.004</td>
<td>.02</td>
<td>-.10</td>
<td>-.01</td>
<td>.06</td>
<td>-.03</td>
<td>-.001</td>
<td>-.69**</td>
<td>.58**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Personal Goal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15. Attributions:</td>
<td>.37**</td>
<td>-.03</td>
<td>-.19*</td>
<td>-.04</td>
<td>.11</td>
<td>-.06</td>
<td>.06</td>
<td>.07</td>
<td>.12</td>
<td>.08</td>
<td>.03</td>
<td>.08</td>
<td>-.02</td>
<td>-.05</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stability</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16. Attributions:</td>
<td>-.07</td>
<td>-.001</td>
<td>.02</td>
<td>.08</td>
<td>-.09</td>
<td>-.07</td>
<td>.02</td>
<td>-.04</td>
<td>.02</td>
<td>.03</td>
<td>-.03</td>
<td>.01</td>
<td>-.07</td>
<td>.05</td>
<td>.05</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Controllability</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17. Attributions:</td>
<td>-.03</td>
<td>.04</td>
<td>-.04</td>
<td>.06</td>
<td>-.05</td>
<td>-.04</td>
<td>.03</td>
<td>-.05</td>
<td>.07</td>
<td>-.06</td>
<td>.03</td>
<td>-.06</td>
<td>.03</td>
<td>.07</td>
<td>14+</td>
<td>.80**</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Internal vs. External</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18. Self-Satisfaction</td>
<td>.76**</td>
<td>.01</td>
<td>-.25**</td>
<td>.004</td>
<td>.26**</td>
<td>-.07</td>
<td>-.14+</td>
<td>.14*</td>
<td>.31**</td>
<td>.36**</td>
<td>.14*</td>
<td>.19**</td>
<td>-.01</td>
<td>.15*</td>
<td>40**</td>
<td>.22**</td>
<td>.13</td>
<td>1</td>
</tr>
</tbody>
</table>

*Note.* *p < .05; **p < .01, +p < .10. Pearson correlations were run for correlations between two continuous variables, point biserial correlations were run for correlations between dichotomous and continuous variables, and phi correlations were run for correlations between two dichotomous variables.
Research Question 1a: Do Students With High Self-Regulation Perform Better Than Students With Low Self-Regulation On A Math Test

To investigate specifically whether high and low self-regulation results in differences in achievement level, individual t-tests were analyzed with high and low self-regulation as the grouping variable and achievement as the dependent variable. Unstandardized scores were analyzed for individual self-regulation processes. Overall, 18 comparisons were conducted. It is important to note that as a result of the high number of statistical comparisons made on the same data, Type I error is inflated. However, the rationale to continue with these analyses was to investigate whether the extreme ends of each self-regulation variable would result in different levels of achievement.

In the t-test analyses, all continuous self-regulation variables (i.e., grade goal, goal difficulty, self-efficacy, math test and math domain task value, metacognitive monitoring [understanding and confidence in answer], self-satisfaction and all composite variables [overall forethought, performance, self-reflection, and self-regulation composites]) were split into high and low groups by a standard deviation split around the mean (i.e., scores that were one standard deviation above the mean coded as high [= “1”] and one standard deviation below the mean as low [= “0”]). Although categorizing a continuous variable is typically advised against given the methodological and empirical flaws it presents (e.g., decreased power and increased Type II error), the rationale was to investigate the extremes ends of each self-regulation process and to see whether that would result in differences in achievement. This process was used to investigate all continuous self-regulation variables. However, for the three category self-evaluation variable, a one-way
ANOVA was run with self-evaluation as the independent variable and achievement on the math test as the dependent variable.

Cohen’s effect size \((d)\) was also calculated for each of the comparisons by taking the mean difference between the two groups (high and low self-regulation) and dividing it by the pooled standard deviation. The goal of these analyses was to investigate whether self-regulation (from its individual components, to different phases, to self-regulation as a whole) as measured by the SRAT:MMA was able to differentiate student achievement levels. The results of all the t-tests including the effect sizes are presented in Table 4.

In terms of the individual t-tests, the results revealed that the direction of the relationship was generally consistent with previous research. The statistically significant results included goal difficulty, self-efficacy, math domain task value, metacognitive monitoring, stable attributions, and self-satisfaction. Specifically, students who reported higher levels of self-efficacy \((M = 3.25, SD = 1.28)\), math domain task value \((M = 3.14, SD = 1.43)\), metacognitive understanding \((M = 3.45, SD = 1.45)\), metacognitive confidence in one’s answer \((M = 3.58, SD = 1.15)\), stable attributions \((M = 4.24, SD = 2.55)\), and satisfaction \((M = 4.30, SD = .88)\) outperformed students who reported lower levels of self-efficacy \((M = 2.15, SD = 1.22; t(131) = 4.99, p < .001, d = .88)\), math domain task value \((M = 2.93, SD = 1.27; t(83) = 2.61, p = .01, d = .72)\), metacognitive understanding \((M = 2.21, SD = 1.28; t(56) = 3.58, p = .001, d = .91)\), metacognitive confidence in one’s answer \((M = 1.93, SD = 1.22; t(58) = 5.39, p < .001, d = 1.39)\), stable attributions \((M = 2.55, SD = 1.30; t(167) = 5.06, p < .001, d = 1.30)\), and satisfaction \((M = 1.50, SD = 1.02; t(97) = 14.06, p < .001, d = 2.94)\). In terms of goal difficulty, lower...
reports of goal difficulty ($M = 3.23, SD = 1.33$) was associated with higher performance levels than higher reports of goal difficulty ($M = 2.11, SD = 1.32; t(58) = -3.22, p = .002, d = -.83$). In terms of the three category self-evaluation variable, the results revealed no statistically significant findings ($F(2, 183) = .46, p = .63, \eta^2_p = .005$), indicating that students achieved at similar levels on the math test regardless of whether they judged the success or failure of their performance based on personal goals, previous performance, or social comparison.

In terms of the differences across phases, the results revealed that overall, students who engaged in more adaptive performance ($M = 3.50, SD = 1.48$) and self-reflection phase processes ($M = 4.65, SD = .49$) achieved higher than students who engaged in less adaptive performance ($M = 1.87, SD = 1.13; t (58) = 4.82, p < .001, d = 1.24$) and self-reflection phase ($M = 2.72, SD = 1.42; t (37) = 5.32, p < .001, d = 1.81$) processes. In terms of the dichotomous overall high/low self-regulation scale that combined all the individual self-regulation variables, the results showed that students with high self-regulation achieved higher ($M = 4.00, SD = 1.20$) than students with low self-regulation ($M = 2.13, SD = 1.39; t (44) = 4.88, p < .001; d = 1.45$).
Table 4

*Individual T-Tests with High/Low Self-Regulation as the Grouping Variable and Achievement as the Independent Variable*

<table>
<thead>
<tr>
<th>Self-Regulation</th>
<th>High Self-Regulation</th>
<th>Low Self-Regulation</th>
<th>T</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Grade Goal</td>
<td>2.78</td>
<td>1.43</td>
<td>2.25</td>
<td>1.00</td>
</tr>
<tr>
<td>Goal Difficulty</td>
<td>2.12</td>
<td>1.32</td>
<td>3.23</td>
<td>1.34</td>
</tr>
<tr>
<td>Process vs Outcome Goal</td>
<td>2.82</td>
<td>1.33</td>
<td>2.58</td>
<td>1.51</td>
</tr>
<tr>
<td>Self-Efficacy</td>
<td>3.25</td>
<td>1.28</td>
<td>2.15</td>
<td>1.22</td>
</tr>
<tr>
<td>Strategic Planning</td>
<td>2.66</td>
<td>1.32</td>
<td>2.79</td>
<td>1.46</td>
</tr>
<tr>
<td>Task Value: Math Test</td>
<td>2.85</td>
<td>1.41</td>
<td>2.93</td>
<td>1.27</td>
</tr>
<tr>
<td>Task Value: Math Domain</td>
<td>3.14</td>
<td>1.43</td>
<td>2.25</td>
<td>1.02</td>
</tr>
<tr>
<td>Metacognition: Understanding of Question</td>
<td>3.45</td>
<td>1.45</td>
<td>2.21</td>
<td>1.28</td>
</tr>
<tr>
<td>Metacognition: Confidence in Answer</td>
<td>3.58</td>
<td>1.15</td>
<td>1.93</td>
<td>1.22</td>
</tr>
<tr>
<td>Attentional Control</td>
<td>2.89</td>
<td>1.49</td>
<td>2.66</td>
<td>1.34</td>
</tr>
<tr>
<td>Attributions: Stability</td>
<td>4.24</td>
<td>1.30</td>
<td>2.55</td>
<td>1.30</td>
</tr>
<tr>
<td>Attributions: Controllability</td>
<td>2.66</td>
<td>1.33</td>
<td>2.91</td>
<td>1.61</td>
</tr>
<tr>
<td>Attributions: Internal vs External</td>
<td>2.72</td>
<td>1.38</td>
<td>2.88</td>
<td>1.55</td>
</tr>
<tr>
<td>Self-Satisfaction</td>
<td>4.30</td>
<td>.88</td>
<td>1.50</td>
<td>1.02</td>
</tr>
<tr>
<td>Overall Forethought Phase Processes</td>
<td>2.89</td>
<td>1.29</td>
<td>2.33</td>
<td>1.15</td>
</tr>
<tr>
<td>Overall Performance Phase Processes</td>
<td>3.50</td>
<td>1.48</td>
<td>1.87</td>
<td>1.13</td>
</tr>
<tr>
<td>Overall Self-Reflection Phase Processes</td>
<td>4.65</td>
<td>.49</td>
<td>2.73</td>
<td>1.42</td>
</tr>
<tr>
<td>Overall Self-Regulation</td>
<td>4.00</td>
<td>1.19</td>
<td>2.13</td>
<td>1.39</td>
</tr>
</tbody>
</table>

*p < .05; **p < .01; ***p < .001; + p < .10
Research Question 1b: Does The Relationship Between The Different Phases Of Self-Regulation Vary Across Levels Of Student Achievement On A Math Test

To assess the extent to which the relationship between the different phases of self-regulation may vary across levels of achievement, a series of moderated multiple regressions were conducted following the procedures outlined by Cohen, Cohen, West and Aiken (2003). Specifically, the relationship between the different phases of self-regulation were expected to vary as a function of achievement, thus, achievement on the math test was used as the moderator variable and as a proxy for both future and prior performance. Although actual prior performance and future performance (e.g., final course grades or final test grades) scores would have been more theoretically and empirically sound, it was not possible to gather this data because students were guaranteed anonymity. First, three interaction terms were created: forethought composite x continuous test grade; performance composite x continuous test grade; and self-reflection composite x continuous test grade. Second, all independent and moderator variables were centered at their means. Finally, three moderated regressions were then performed by entering student gender and minority status as controls in the first step, followed by the individual achievement and self-regulation phase main effects in the second step, and the relevant cross-product interaction term in the last step to examine whether the level of achievement moderated the relationship between each of the self-regulation phases. The three regressions were:

1) Performance composite variable as the dependent variable:
   - Step 1 = Gender and minority status
- Step 2 = Grade and forethought phase composite
- Step 3 = Grade x forethought phase interaction

2) Self-reflection variable as dependent variable
- Step 1 = Gender and minority status
- Step 2 = Grade and forethought phase composite
- Step 3 = Grade x forethought phase interaction

3) Self-Reflection composite variable as dependent variable
- Step 1 = Gender and minority status
- Step 2 = Grade and performance phase composite
- Step 3 = Grade x performance phase interaction

Significant interactions were then plotted using simple slope techniques following the guidelines from Cohen et al., (2003). Specifically, Cohen et al. (2003) suggests plotting three equations along x and y to interpret a significant interaction: one equation representing one standard deviation above the mean of the moderator, one equation representing one standard deviation below the mean of the moderator, and one equation representing the mean of the moderator.

The results revealed two significant interactions between: 1) the forethought phase and self-reflection phase (interaction term $t = 3.34, p < .001$); and 2) the performance phase and self-reflection phase (interaction term $t = 2.01, p < .05$). In terms of the interaction between forethought and self-reflection, the results revealed that after controlling for demographics and individual main effects, a significant change in variance was accounted for when the interaction term was added to the model ($\Delta R^2 = .06; R^2 = .21$,}
Specifically, the interaction term accounted for 6% of the variance in the self-reflection phase. This suggests that the relationship between the forethought phase and self-reflection phase varied across levels of student achievement. An interpretation of the simple slope plots (Figure 2) showed that there was a positive relationship between the forethought and self-reflection phase processes for students who achieved high. However, there was a negative relationship between the forethought and self-reflection phase processes for students who achieved low.

*Figure 2.* Graph of the interaction between forethought and self-reflection phase composite variables by math test achievement levels.

In terms of the interaction between performance and self-reflection, the results revealed that after controlling for demographics and individual main effects, a significant
change in variance was accounted for when the interaction term was added to the model ($\Delta R^2 = .02; \, R^2 = .18, \, F(5, \, 141) = 6.19, \, p < .001$). Specifically, the interaction term accounted for 2% of the variance in the self-reflection phase. This result indicates that the relationship between the performance and self-reflection phase processes varied across levels of student achievement. An investigation of the simple slopes plot (Figure 3) indicated that there was a positive relationship between the self-reflection and performance phase processes in students who achieved high. However, in students who achieved low, there was a negative relationship between the self-reflection and performance phase processes.

Figure 3. Graph of the relationship between performance and self-reflection phase composite variables by math test achievement levels.
No significant interaction was found in the relationship between the forethought and performance phase. Specifically, after controlling for gender and minority status in the first step, and then the individual effects in the second step, the performance by forethought interaction did not a statistically significant increase in $R^2 (\Delta R^2 < .001; R^2 = .21; t = -.21, p = .83)$ with a statistically significant ANOVA, $(F(5, 159) = 8.5, p < .001)$. This may indicate that overall, student achievement and engagement in forethought phase processes are related to performance phase processes, but this relationship is not moderated by high and low achievement differences on the math test.

**Research Question 1c: Can The Self-Regulation Variables, As Measured By The SRAT: MMA, Predict College Student Achievement On A Math Test**

To address this research question a hierarchical regression was employed. Prior to conducting the regression, several assumptions of multiple regression as suggested by Cohen et al. (2003) were investigated. First, collinearity statistics (i.e., VIF and tolerance statistics) revealed acceptable levels of multicollinearity. Second, P-P plots with all the continuous variables showed that all the individual continuous self-regulation variables were normally distributed. Third, homoscedasticity was investigated by plotting the regression standardized residuals against the standardized predicted value. A visualization of the plots revealed that the assumption of homoscedasticity was met. Finally, the last assumption tested was the normality of residuals, or that the errors were normally distributed. To test this assumption, a visual inspection of a histogram with a superimposed normal curve was conducted with each individual standardized residual.
values of each predictor. Overall, the normality of residuals was met for all self-regulation predictors.

After the assumptions of regression were tested, all the continuous self-regulation variables were then mean centered. The three level categorical self-evaluation variable was separated and dummy coded into three individual variables: personal goal (coded as Yes = 1, No = 0), prior performance (coded as Yes = 1, No = 0), and social comparison (coded as Yes = 1, No = 0). The prior performance and social comparison dummy variables were entered into the regression, thus using personal goal as the reference group. The demographic variables were entered into the first step of the regression as controls, followed by the self-regulation variables in the forethought phase in the second step, the self-regulation variables in the performance phase in the third step, and finally the self-regulation variables in the self-reflection phase in the fourth step. Overall, the purpose of this analysis was to investigate how self-regulation, when treated as a three phased model, was related to student achievement on the math test. In addition, the main goal of this analysis was to investigate the predictive validity of the SRAT:MMA. See Table 5 for the regression results.
Table 5

Hierarchical Regression of all Individual Self-Regulation Variables entered into three Models Predicting Student Math Achievement

<table>
<thead>
<tr>
<th>Variables</th>
<th>R²</th>
<th>R² Change</th>
<th>F</th>
<th>Sig. F Change</th>
<th>β</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1: Controls</td>
<td>.02</td>
<td>.02</td>
<td>1.40</td>
<td>.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minority Status</td>
<td>.11</td>
<td>1.35</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>-.08</td>
<td>-.93</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 2: Forethought Phase</td>
<td>.21</td>
<td>.19</td>
<td>3.84</td>
<td>&lt;.001</td>
<td>.11</td>
<td>1.37</td>
</tr>
<tr>
<td>Minority Status</td>
<td>.11</td>
<td>1.37</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>-.07</td>
<td>-.87</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade Goal</td>
<td>-.01</td>
<td>-.10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Goal Difficulty</td>
<td>-.17</td>
<td>-1.82+</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Process vs. Outcome Goals</td>
<td>.06</td>
<td>.69</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-Efficacy</td>
<td>.21</td>
<td>2.19*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strategic Planning</td>
<td>-.04</td>
<td>-.57</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task Value: Math Test</td>
<td>-.07</td>
<td>-.80</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task Value: Math Domain</td>
<td>.19</td>
<td>2.16*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 3: Performance Phase</td>
<td>.33</td>
<td>.12</td>
<td>5.30</td>
<td>&lt;.001</td>
<td>.15</td>
<td>1.97+</td>
</tr>
<tr>
<td>Minority Status</td>
<td>.15</td>
<td>1.97+</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>-.04</td>
<td>-.50</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade Goal</td>
<td>-.09</td>
<td>-1.11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Goal Difficulty</td>
<td>-.06</td>
<td>-.63</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Process vs. Outcome Goals</td>
<td>.08</td>
<td>1.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-Efficacy</td>
<td>.10</td>
<td>1.11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strategic Planning</td>
<td>-.05</td>
<td>-.73</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task Value: Math Test</td>
<td>-.06</td>
<td>-.78</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task Value: Math Domain</td>
<td>.16</td>
<td>1.99*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metacognition: Understanding of Question</td>
<td>-.03</td>
<td>-.29</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metacognition: Confidence in Answer</td>
<td>.43</td>
<td>3.71***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attentional Control</td>
<td>.16</td>
<td>2.18*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 4: Self-Reflection Phase</td>
<td>.72</td>
<td>.39</td>
<td>17.23</td>
<td>&lt;.001</td>
<td>.07</td>
<td>1.37</td>
</tr>
<tr>
<td>Minority Status</td>
<td>.07</td>
<td>1.37</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>-.04</td>
<td>-.76</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade Goal</td>
<td>-.03</td>
<td>.45</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Goal Difficulty</td>
<td>-.02</td>
<td>-.25</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Process vs. Outcome Goals</td>
<td>.10</td>
<td>1.94+</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-Efficacy</td>
<td>.07</td>
<td>1.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strategic Planning</td>
<td>-.03</td>
<td>-.68</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task Value: Math Test</td>
<td>.06</td>
<td>1.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task Value: Math Domain</td>
<td>.04</td>
<td>.64</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metacognition: Understanding of Question</td>
<td>-.08</td>
<td>-1.08</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metacognition: Confidence in Answer</td>
<td>.19</td>
<td>2.34*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attential Control</td>
<td>.02</td>
<td>.46</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-Evaluation: Prior Performance</td>
<td>.01</td>
<td>.19</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-Evaluation: Social Comparison</td>
<td>-.09</td>
<td>-1.70+</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attributions: Stability</td>
<td>.04</td>
<td>.71</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attributions: Controllability</td>
<td>.11</td>
<td>1.16</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attributions: Internal vs. External</td>
<td>-.03</td>
<td>-.32</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-Satisfaction</td>
<td>.74</td>
<td>11.55***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p < .05; **p < .01; ***p < .001; + p < .10
The results revealed that all three (forethought, performance, and self-reflection) self-regulation phase processes played a significant role in predicting student achievement. Results suggested that first model, minority status and gender were not statistically significant predictors of student math test achievement. However, results revealed that the forethought phase variables accounted for 21% of the variance in student math test achievement ($\Delta R^2 = .19; R^2 = .21; F(9, 132) = 3.84, p < .001$). A significant change was detected in accounted variance with the addition of the third model with the performance phase variables ($\Delta R^2 = .12; R^2 = .33; F(12, 129) = 5.29, p < .001$). Finally, in the fourth regression model, with the self-reflection phase variables added to the previous predictors, a significant change was detected ($\Delta R^2 = .39; R^2 = .72; F(18, 123) = 17.23, p < .001$). All together, the self-regulation variables accounted for 70% of the variance in student achievement on the math test.

The regression coefficients under the forethought phase model revealed that self-efficacy ($\beta = .21, t = 2.19, p < .05$) and task value in the math domain ($\beta = .19, t = 2.16, p < .05$) were statistically significant predictors in student achievement on the math test. In terms of the performance phase, having confidence in one’s answer (metacognition) ($\beta = .43, t = 3.71, p < .001$), and use of attentional control strategies ($\beta = .16, t = 2.18, p < .05$) were statistically significant predictors in student math achievement. Task value in the math domain continued to predict student achievement ($\beta = .16, t = 1.99, p < .05$) whereas self-efficacy became less predictive when the performance phase variables were added into the model. Finally, under the self-reflection phase, self-satisfaction ($\beta = .74, t = 2.16, p < .05$) was a significant and positive main effect on student achievement on the
math test and confidence in one’s answer (metacognition, \( \beta = .19, t = 2.34, p < .05 \)) from the performance phase continued to statistically significantly predict student achievement.

**Research Question 2: Self-Regulation And Stereotype Threat**

Fidelity analyses revealed that the stereotype threat manipulation failed to evoke stereotype threat in this sample of students. The means and standard deviations by stereotype threat condition and self-regulation are reported in Appendix F. This study attempted to mirror the stereotype threat manipulation strategy in Good et al. (2008) for two main reasons: 1) the natural context in which stereotype threat was invoked (i.e., in a math classroom as opposed to a laboratory setting); and 2) the method that stereotype threat was manipulated, in that it was closer to the reality of everyday classrooms than more traditional methods of manipulating stereotype threat. Good et al. (2008) recognized the practical limitations of stereotype threat results due to the heavy reliance on highly controlled laboratory settings. Therefore, one of the main goals in her study was to show that stereotype threat does occur in the regular classroom.

Although the present study attempted to follow the procedures in Good et al. (2008) there were several unique methodological differences (see Table 6), which are listed below. The section below will describe the similarities and differences between Good et al. (2008) and this present study. This was done in an effort to determine the reason why the manipulation succeeded in Good et al. (2008), but failed in this present study.
Table 6

*Methodological Similarities and Differences between Good et al. (2008) and this Present Study*

<table>
<thead>
<tr>
<th>Method</th>
<th>Methodological Similarities and Differences</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Study Design</strong></td>
<td>Good et al. (2008) 2 (gender) X 2 (high low stereotype threat)</td>
</tr>
<tr>
<td><strong>Sample</strong></td>
<td>This Present Study Identical to Good et al. (2008)</td>
</tr>
<tr>
<td><strong>Size (N)</strong></td>
<td>157                                                           191</td>
</tr>
<tr>
<td><strong>Gender Breakdown</strong></td>
<td>100 Males; 57 Females                                      111 Males; 80 Females</td>
</tr>
<tr>
<td><strong>Context</strong></td>
<td>One rigorous college calculus course, third course in a three course calculus sequence</td>
</tr>
<tr>
<td></td>
<td>Four college math courses across five classes: Two upper level courses; two second courses in a three course calculus sequence; one third course in a three course calculus sequence</td>
</tr>
<tr>
<td><strong>Measures</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Achievement</strong></td>
<td>One researcher developed test from GRE math questions, unknown number of questions</td>
</tr>
<tr>
<td></td>
<td>Five different tests created by the professor based on final course exam, five questions</td>
</tr>
<tr>
<td><strong>Timed Test</strong></td>
<td>20 minutes to finish test                                   25-30 minutes to finish test, based on professor requests</td>
</tr>
<tr>
<td><strong>Survey Packet</strong></td>
<td>Individual sections sealed                                  Individual sections not sealed</td>
</tr>
<tr>
<td><strong>Survey Packet</strong></td>
<td>Informed consent, high low stereotype threat manipulation, pretest questionnaire, math test, posttest questionnaire, demographic questionnaire, debriefing form</td>
</tr>
<tr>
<td><strong>Organization</strong></td>
<td>Identical to Good et al. (2008)</td>
</tr>
<tr>
<td><strong>Pilot Testing</strong></td>
<td>Achievement test was piloted with 12 calculus students to determine the degree of</td>
</tr>
<tr>
<td></td>
<td>Achievement test was not piloted. Professors were instructed to create a practice</td>
</tr>
<tr>
<td>Method</td>
<td>Good et al. (2008)</td>
</tr>
<tr>
<td>---------------------------------------------</td>
<td>--------------------------------------------------------</td>
</tr>
<tr>
<td><strong>Methodological Similarities and Differences</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Difficulty</strong></td>
<td>test similar to the final test. Tests were given to two math experts who determined that all tests were equal in difficulty levels.</td>
</tr>
<tr>
<td><strong>Procedure</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Student Knowledge about the Practice Test</strong></td>
<td>Students were told one week prior to the test that there would be a practice test for the upcoming final.</td>
</tr>
<tr>
<td><strong>Test Description</strong></td>
<td>Test described as extra credit practice test for upcoming final</td>
</tr>
<tr>
<td><strong>Extra Credit</strong></td>
<td>Amount of extra credit rewarded would be dependent on performance on math test</td>
</tr>
<tr>
<td><strong>Experiment Context</strong></td>
<td>Test administered in classroom conditions</td>
</tr>
<tr>
<td><strong>Experiment Timing</strong></td>
<td>Test administered during university reading week</td>
</tr>
<tr>
<td><strong>Experimental Controls</strong></td>
<td>Experimenter was knowledgeable about the hypothesis, but blind to the condition</td>
</tr>
<tr>
<td><strong>High Stereotype Threat Manipulation</strong></td>
<td>Informed students test was diagnostic ability</td>
</tr>
<tr>
<td><strong>Low Stereotype Threat Manipulation</strong></td>
<td>Maintained diagnostic ability of the test, added statement regarding no gender differences</td>
</tr>
<tr>
<td><strong>Fidelity Analysis</strong></td>
<td>2 X 2 ANOVA: High Low stereotype threat condition X Gender on test bias perceptions</td>
</tr>
</tbody>
</table>
As reported in Table 6, this present study attempted to mimic as closely as possible the experimental procedures in Good et al. (2008). The overall design of this study was identical Good et al. (2008) in terms of the procedures and stereotype threat manipulation. This study differed with Good et al. (2008) in the context and measures. In the Good et al. (2008) study, students from one course participated in the study whereas in this present study, students were from one of five individual courses. Another point of distinction was in the achievement measure: questions were taken from the GRE test in Good et al. (2008) whereas this study used professor created practice tests.

Good et al. (2008) did not provide a detailed description of the math test used in their study to investigate stereotype threat (e.g., no information about the number of questions, or the actual test itself). Therefore, a decision was made to allow professors to create their own practice test for an upcoming final, given specific parameters such as number of questions. Efforts were also made in this study to ensure that all students were given approximately the same amount of time to complete the test. However, some professors felt strongly about providing their students with a full 30 minutes to complete the test whereas others felt that 25 minutes was appropriate.

Ultimately, individual professor requests to provide either 30 or 25 minutes to complete the test were accepted. This was done to assess self-regulation and stereotype threat in a naturalistic context, using an achievement measure that was as closely aligned to the everyday math environment as possible. Unfortunately, the fidelity analyses in this present study still failed to confirm the stereotype threat manipulation. These results will be further discussed later in Chapter 5.
Chapter Five

The major purpose of the study was to investigate the initial validity of the SRAT:MMA. The secondary purpose was to examine whether self-regulation can alleviate the effects of stereotype threat on women enrolled in undergraduate math classes. The classic microanalytic approach was modified to assess student self-regulation with a self-report survey and stereotype threat was manipulated through test instructions. Generally, the initial evidence regarding the validity of the modified microanalytical approach suggested that the SRAT:MMA has the potential to be a practical and valid alternative current self-regulation assessment techniques. The following sections will first discuss the findings of the first three research sub-questions including whether the hypotheses were confirmed and how they align or diverge with prior research. Second, the stereotype threat manipulation will be discussed, including an explanation on the possible reasons why the manipulation failed to invoke stereotype threat. Finally, this chapter will conclude with an overall discussion of the practical implications, limitations, and suggestions for future research.

The first research sub-question was: Do students with high self-regulation perform better than students with low self-regulation on a math test? It was expected that students who report higher levels of self-regulation practices will achieve higher than students who report lower levels of self-regulation practices. This hypothesis was generally supported. Specifically, the results from t-test analyses revealed that students who reported lower levels of goal difficulty, and higher levels of self-efficacy, math domain identification, metacognitive monitoring, stable attributions, and self-satisfaction achieved higher on the math test than students who reported higher levels of goal difficulty, and lower levels of self-efficacy, math domain identification, metacognitive monitoring, stable attributions, and self-satisfaction. Furthermore, students who reported stronger engagement in overall performance and self-reflection phase processes achieved higher than students who reported less engagement in overall performance and self-reflection phase processes. Finally, in terms of overall self-regulation, students with stronger overall self-regulation achieved higher than students with weaker overall self-regulation.

These results regarding self-regulation and achievement are generally consistent with prior research in that students who are skilled in self-regulation tend to achieve higher academically than students who are less skilled in self-regulation (DiBenedetto & Zimmerman, 2010, 2013; Kitsantas, 2002; Schunk et al., 2008; Zimmerman, 2008a, 2010, 2013).
2008b). However, it was interesting that out of all three dimensions of attributions (stability, controllability, and locus of control), students who made more stable attributions were more likely to achieve higher than students who made less stable attributions.

Prior research (Weiner, 1986) suggests that students who adopt attributions that change over time will engage in behaviors that increase the chances of high achievement more so than students adopt attributions that do not change over time. If attributions are stable, then it suggests that performance is unlikely to change over time. For example, a student could attribute poor performance to: “I’m just not good at math,” which would mean that there is an expectation for underperformance. This attribution would be considered as stable (lack of ability does not change over time), internal (lack of ability comes from within, and not from external or environmental factors) and uncontrollable (there is not much that can be done to change lack of ability [Schunk et al., 2008]). Conversely, another student could attribute high performance to: “I’m just a math person,” which is also a stable, internal, and uncontrollable attribution, but there is an expectation for high performance. Therefore, both attributions for success and failure could fall along similar dimensions of attributions, but the key difference here is the expectation for success. In fact, Schunk et al. (2008) suggest that the stability dimension is closely related to expectancy for success.

Therefore, similar to the findings from Gibb et al., (2002) attributions have different consequences for different groups of students depending on the expectancy for success. In this study, students were enrolled in an upper level math class, or in a second
or third calculus sequence course. All of the courses that students were enrolled in had at least one perquisite. As a result, all of the student participants must have had some successful experiences in math in order to be enrolled in the courses that they were in, which would explain why these students were not affected by typically maladaptive attributional styles (Gibb et al., 2002).

It is important to note that although grade goal and overall engagement in forethought phase processes did not significantly differentiate achievement, the effect sizes for grade goal and overall forethought phase were moderate ($d = .43$ and $.46$, respectively), suggesting that there may be a moderate association between grade goal and overall forethought phase processes with achievement on the math test. However, this should be interpreted with caution given the lack of statistical significance. Process outcome goals, strategic planning, math test value, attentional control, controllable and internal/external attributions (i.e., locus of control), and self-evaluation, however, yielded rather weak effect sizes ($d = .17 - .06$) and did not statistically significantly differentiate high and low student achievement. This finding diverges with prior research which suggests that process goals are generally more important than outcome goals (Zimmerman, 2008b), and that strategic planning is an integral process that guides behaviors and approaches to achieve a certain goal (Winne, 1997; Zimmerman, 2008a). However, it is important to take into account the context in which the task occurred in. The math test was an optional extra credit test that carried no negative consequences if the students performed poorly. Prior research has suggested that students tend to engage in more self-regulatory behavior for consequential tasks than for non-consequential tasks.
A possible indication of student perceptions about how seriously they took the test was with math task value. Specifically, the self-regulatory patterns suggested that task value for the math domain was more predictive of student achievement on the math test than task value for the math test. Perhaps the non-consequential nature of the test influenced the lack of general student engagement in some of the self-regulatory behaviors. Future research should investigate the relationships between different self-regulatory behaviors within a meaningful, consequential context.

Although not all of the individual self-regulation variables succeeded in differentiating high versus low student achievement, there was evidence to support that overall, general engagement in self-regulation is associated with higher test scores. Specifically, student engagement in performance and self-reflection phase processes was significantly associated with higher test scores whereas student engagement in forethought phase processes approached statistical significance in differentiating high and low achievement. These findings compliment DiBenedetto and Zimmerman (2010, 2013) and Kitsantas (2002), who found that students with stronger overall self-regulation profiles achieved higher than students who had weaker overall self-regulation profiles. These findings also extend Kitsantas (2002) in that students who engage in stronger performance and self-reflection phase processes overall achieved higher than students with weaker engagement in the performance and self-reflection processes.

Generally, the findings of this first research question support prior research in that self-regulation overall is an important determinant in student achievement (Zimmerman, 2000, 2008a). These results also support Zimmerman’s (2000) three phased model of
self-regulation, where the different phases of self-regulation determine student levels of achievement (DiBenedetto & Zimmerman, 2010, 2013). Finally, the results with regard to stable attributions being indicative of higher performance also contribute to the notion that self-regulation is a context specific event (Cleary, 2011; DiBenedetto & Zimmerman, 2010, 2013; Winne & Perry, 2000; Zimmerman 2008a). These results supplement the next finding regarding the moderating role of achievement on the relationship between the different phases.

Finding Two: The Relationship between the 1) Forethought Phase and Self-Reflection Phase and the 2) Performance and Self-Reflection Phase Varies Across Levels of Student Achievement.

The second research sub-question was: Does the relationship between the different phases of self-regulation vary across levels of student achievement on a math test? It was hypothesized that the relationship between the different phases of self-regulation will be positive, or more adaptive, for students who achieve high than for students who achieve low. Specifically, all three phases of self-regulation were expected to be positively related to one another in students who achieve high (Zimmerman, 2000, 2008a). The results partially supported this hypothesis. Specifically, the results showed two significant interactions. First, there was a positive relationship between the forethought and self-reflection phase in students who achieved high on the test whereas there was a negative relationship between these phases in students who achieved low on the test. Second, there was a positive relationship between the performance and self-
reflection phase in students who achieved high, but this relationship was reversed for students who achieved low.

According to Zimmerman (2000, 2008a), the most adaptive form of self-regulation is when the phases are positively and cyclically interrelated. These findings confirm Zimmerman’s notion in that students who achieved high exhibited a positive relationship between the different phases whereas students who achieved low did not. The significant interaction between the forethought and self-reflection phase may indicate that students who did not perform high on the math test tended to overestimate their abilities during the forethought phase, which generated a negative response in the self-reflection phase. Student self-satisfaction would be negatively impacted when students who expected to achieve high on the test actually received a low grade and positively impacted when students who expected to achieve high actually received a high grade. In terms of the significant interaction between the performance and self-reflection phase, the same interpretation could be applied. Specifically, students who achieved low expected to achieve high because they had engaged in positive metacognitive monitoring and attention focusing strategies during the performance phase, but their low performance outcomes were at odds with their positive behaviors in the performance phase, generating a negative relationship. However, this relationship would be reversed for students who achieved high on the test, where they had engaged in positive performance phase processes which resulted in high self-reflection phase processes.

These data may suggest that low achieving students have difficulty judging their capabilities. Perhaps these students begin with a positive sense of self-efficacy and set
high goals, however, because they had overestimated their actual ability, when they attempted the task and realized that they have yet to gain the skills to do well, they become dissatisfied and adopt more maladaptive attributions (DiBenedetto & Zimmerman, 2013). The results would be opposite for those who had achieved high, where their positive motivation in the forethought phase aligned with their performance and thus positive self-reactions in the self-reflection phase. Similar results have been found in the past, where students with high achievement tended to exhibit more adaptive patterns of self-regulated learning (DiBenedetto & Zimmerman, 2010, 2013; Kitsantas, 2002).

Interestingly, the relationship between the forethought and performance phase did not vary across levels of achievement. This may again point to the notion that all students, regardless of their actual ability, felt that they could do well on the test. As a result of these high beliefs, all students focused on the test and engaged in positive metacognitive monitoring and attention focusing strategies during the performance phase. Although Zimmerman (2000) suggests that high achieving students tend to exhibit stronger motivational beliefs which lead to stronger performance phase behaviors, the characteristics of the sample may explain why it did not confirm Zimmerman’s (2000) notions. Specifically, all students must have had some sort of success in previous math courses, given that the courses they were enrolled in all had at least one prerequisite math course. In fact, the sample statistics showed that all students (except for the ten students who did not report the number of math classes completed) had completed at least one math course prior to enrolling in their current course. Therefore, it may have been that all
of the students in this sample were capable of doing well in math, which would explain
the positive relationship between forethought and performance phase processes
regardless of achievement level. Ultimately, differences emerged during the self-
reflection phase, when students were given their performance outcomes. The negative
reactions from students who achieved low would certainly be negatively related to their
strong forethought phase beliefs and performance phase behaviors.

The results of these moderation analyses may suggest that the self-reflection
phase provides crucial cyclic information to the forethought and performance phases
(DiBenedetto & Zimmerman, 2010). These moderation findings confirm the dynamic
nature of self-regulation and the importance of taking the context into account (Cleary,
2011; DiBenedetto & Zimmerman, 2013, 2010; Zimmerman, 2008a). Overall, this is the
first study, to my knowledge, to assess self-regulation as a set of individual processes, a
composite of processes across phases, and a composite of overall self-regulation. These
first two findings regarding the achievement differences between skilled and less skilled
self-regulation as well as the moderation findings showed that the SRAT:MMA is able to
detect self-regulation deficiencies at the individual process level to the phased and overall
level. The findings support both theoretical (Zimmerman, 2000, 2008a) and empirical
(DiBenedetto & Zimmerman, 2010, 2013) research regarding the nature of self-regulation
and achievement. The following section will now describe the findings regarding the
predictive validity of the SRAT:MMA.
Finding Three: The SRAT:MMA Scales Altogether Explained 70% of the Variance in Student Math Test Scores.

The third research subquestion was: “Can the self-regulation variables, as measured by the SRAT:MMA predict college student achievement on a math test?” It was hypothesized that the self-regulation variables as measured by the SRAT:MMA can indeed predict student math test scores. This hypothesis was supported. Specifically, the results revealed that overall, all of the self-regulation variables, excluding the gender and minority status controls, explained approximately 70% of the variance in student math test scores. This finding aligns with DiBenedetto and Zimmerman, (2013) who found that all of the microanalytically measured self-regulation variables explained 80% of the variance in a student conceptual understanding test.

This finding also aligns with the notion that, in general, microanalytic measures are more predictive of student performance than global self-regulation measures (Cleary, 2011). For example, Kitsantas, Winsler, and Huie (2008) found that global items from the MSLQ accounted for 11% and 8% of first year and third year overall GPA, respectively. Zimmerman and Kitsantas (2014) found that global self-regulation measures (composite of several different global self-regulation scales, including the MSLQ) accounted for 34% of the variance in student GPA scores and further, Cleary and Callan (2014) found that global self-regulation scales accounted for approximately 28% of the variance in student test scores. Cleary and Callan’s (2014) results aligned with Kitsantas (2002), who also found that 24% of the variance in student test scores could be explained by differences in self-regulation. Although Kitsantas (2002) did not utilize the classic microanalytic
approach, it did incorporate some microanalytic methods which included administering forethought scales right before testing. In this present study, by incorporating all critical aspects of self-regulation microanalysis except for individual assessment, 70% of the variance in student achievement can be attributed to self-regulation differences, which is marginally less than the 80% of variance explained by the full microanalytical approach (DiBenedetto & Zimmerman, 2013), but considerably more than the approximately 8% to 34% of variance explained by global measures and partial microanalytic approaches (Cleary & Callan, 2014; Kitsantas 2002; Kitsantas et al., 2008; Zimmerman & Kitsantas, 2014).

The results also revealed that each individual phase of self-regulation contributed unique variance in predicting student math test achievement, mimicking the results of DiBenedetto and Zimmerman (2013). It is also important to note that in this present study, the largest increase in percent of variance explained was from the performance to self-reflection phase, where the self-reflection phase variables added an additional 39% of variance explained above and beyond performance phase variables. In this last model, process goals and metacognition were also important predictors of student math test performance, but the strongest predictor was self-satisfaction. This may indicate that self-satisfaction is an important process within the three phased model that influences other aspects of self-regulation (DiBenedetto & Zimmerman, 2013).

There was an interesting pattern with self-satisfaction, where it was the most important predictor of student math test performance. As previously mentioned, this may potentially be due to the unrealistic goals that students who achieved low set during the
forethought phase. Specifically, in the descriptive statistics of high and low student achievement by self-regulation process (see Table 2), high and low achieving students set approximately the same grade goal (High achieving students grade goal $M = 4.64$; Low achieving students grade goal $M = 4.55$), but clearly achieved at different levels. Evidently, this may have influenced the significant differences in satisfaction levels between those who reported a high grade goal and actually received a high score and those who reported a high grade goal but actually received a low score. This finding leads to the conclusion that students who are not skilled in self-regulation tend to overestimate their abilities, whereas students who are skilled in self-regulation accurately estimate their abilities. Furthermore, this may also indicate that self-satisfaction is highly influenced by other self-regulation processes, or vice versa, where self-satisfaction highly influences other self-regulation processes. Specifically, in the intercorrelations (Table 3), self-satisfaction was positively related to several aspects of self-regulation such as self-efficacy, metacognition, and attributions. In fact, Zimmerman (2011) suggests that self-satisfaction is linked to the goals and strategies that students adopt in the forethought and performance phase as well as the types of achievement attributions students make (Weiner, 1986).

Overall, the initial evidence of validity regarding the SRAT:MMA showed that this modified microanalytic method may be a promising alternative to traditional survey measures, lending support to the first research question. Specifically, the initial evidence of validity showed that: 1) high and low self-regulation levels (both overall self-regulation and individual self-regulation process) differentiated high and low
achievement levels; 2) students who achieved high on the math test tended to exhibit more positive self-regulation patterns consistent with Zimmerman’s (2000) three phased model than students who achieved low on the math test; and 3) the self-regulation processes, as measured by the SRAT:MMA, explained 70% of the variance in student math achievement, demonstrating stronger predictive validity than traditional survey measures. All three results support the first overall research question, which was: Can the SRAT:MMA diagnose college student self-regulated learning at the classroom level during test taking in a math course. Since the stereotype threat manipulation failed in this study, it was not possible to examine whether the self-regulation processes, as measured by the SRAT:MMA, was able to alleviate stereotype threat effects on achievement. This next section will describe the experimental nuances of manipulating stereotype threat.

Stereotype Threat Manipulation

In terms of stereotype threat, it was not possible to investigate this variable because fidelity analyses showed that the stereotype threat manipulation failed. There is evidence to suggest that manipulating stereotype threat is highly nuanced in laboratory settings, and even more so in naturalistic settings (Cromley et al., 2013; Nguyen & Ryan, 2008). In fact, in a review of stereotype threat and math performance, Stoet and Geary (2012) found that 55% of 20 studies had fully replicated the original gender by stereotype threat effect findings from Spencer et al. (1999) (i.e., with a similar sample and study design). However, it is important to note that although the other studies did not fully replicate Spencer et al. (1999), they did detect stereotype threat effects after controlling for a third variable such as gender identity.
Even when the fidelity analysis indicated that the stereotype manipulation had succeeded, as was the case in Good et al., (2008) she and her colleagues were still unable to statistically detect a gender by stereotype threat interaction, which is the critical stereotype threat test. As a result, Good et al. (2008) only included Caucasian students in her study, with the rationale that “These differences might not have yielded results that attained the level of significance because sex differences in math performance may exist primarily among Anglo-American students” (p. 23). This may suggest that even when a stereotype threat manipulation had succeeded, the stereotype threat effect on achievement continues to be difficult to detect.

There are researchers who argue that stereotype threat may have a very small, rather negligible effect on student achievement in natural settings, and that researchers should focus on other variables to explain student achievement differences, especially in the STEM (Science, Technology, Engineering, and Math) areas (Cromley et al., 2013; Stoet & Geary, 2012). Specifically, Cromley et al. (2013) conducted a large scale study (N = 1,358 college students) investigating longitudinal perceptions of racial and gender stereotype bias (i.e., defined as the degree to which individuals believe they will be stereotyped based on their gender or ethnicity), and their effects on achievement and retention in STEM majors. The assumption here was that stereotype bias is an indicator of stereotype threat, where students who have high perceptions of stereotype bias are more vulnerable to stereotype threat than students who have low perceptions of stereotype bias. Cromley et al. (2013) found that: 1) gender and racial stereotype bias was not strongly related to course grades; 2) although stereotype bias perceptions generally
increased over time, it also decreased between breaks and semesters; and 3) stereotype bias was not a strong predictor of retention in STEM majors. Overall, less than 3% of the variance in course grades or STEM retention could be explained by racial or gender stereotype bias. As a result, Cromley et al. (2013) suggest that more attention should be placed on other variables, besides stereotype threat, to explain the racial/gender achievement and retention gaps in STEM areas.

Overall, prior research suggests that stereotype threat is generally difficult to manipulate and detect, especially in naturalistic settings. On top of these inherent limitations to manipulating stereotype threat, there are two specific potential explanations as to why the stereotype threat manipulation did not work in this study: level of math class and non-consequential nature of the test. First, stereotype threat is most likely to occur when a valued ability is questioned by a social stereotype Therefore, this study attempted to collect data from upper level or non-introductory math courses, where the students enrolled are more likely to value their math skills than students enrolled in a lower level math course. However, in this study, only two of the five classes were truly upper level courses: Math 351: Probability and Math 352: Statistics. During participant recruitment, only two upper level math professors were willing to allow this research to be conducted in their classroom due to the large amount of material covered within the semester. Additionally, the student sample in these classrooms was rather low (22 students in Math 351 and 10 students in Math 352), making up only 16% of the total sample.
The second reason why the stereotype manipulation failed may be due to the lack of consequences on the test. Ideally, to align with the natural context of everyday classrooms, the math test should have been described to students as a graded test and administered after the end of a unit. However, professors had concerns about deceiving their students and were not willing to provide class time for data collection. Therefore, the test was administered as an opportunity for extra credit during a study session outside of class during university reading week, and that the amount of extra credit received would be contingent upon performance. Although Good et al. (2008) proved that stereotype threat can be detected in an extra credit practice test, perhaps it was the combined effect of both value and lack of consequences that led to the failed stereotype threat manipulation. The results showed that students generally did not care about their performance on the math test as much as their performance in math overall. This result, coupled with the lack of real consequences, may have led to the inability to confirm the stereotype threat manipulation.

Overall, the results from previous stereotype threat studies and the failed stereotype threat manipulation in this study may point to an important question: how exactly does stereotype threat operate in the natural environment? Although Cromley et al. (2013) and Stoet and Geary (2012) suggested that stereotype threat may not play as strong of a role in student achievement and retention as we thought, they continue to advocate that stereotype threat exist and it does influence student behaviors and achievement. The results from Cromley et al. (2013) further confirmed a meta-analysis conducted by Nguyen and Ryan (2008), where researchers found that although stereotype
threat research is highly robust, the overall effect size across studies was not strong \((d = .26\) across race and gender stereotype threat; \(d = .32\) for race stereotype threat; \(d = .25\) for gender stereotype threat), leading to the conclusion that although stereotype threat does indeed exist, it operates differently under different conditions. Although it is clear that stereotype threat is difficult to invoke and detect, it still influences student achievement. For that reason, stereotype threat is a topic that researchers should continue to investigate.

**Practical Implications, Limitations, and Suggestions for Future Research**

Overall, the results revealed that the SRAT:MMA is a promising alternative to current self-regulation assessment techniques. The initial findings of this modified microanalytical approach showed initial evidence of predictive validity. Self-regulation researchers have advocated for a more context specific, meaningful approach to assessing self-regulation as a set of dynamic processes (Cleary et al., 2008, Kitsantas & Zimmerman, 2002; Winne, 2000; Zimmerman, 2008a). However, current methodologies that take into account the context and dynamic nature of self-regulation are time-consuming and inefficient for both researchers and teachers. This modified microanalytic approach attempted to balance both the theoretical soundness of self-regulation assessment as well as the practical and logistical needs of researchers and classroom teachers/professors. The following sections will first describe the practical implications of using the SRAT:MMA and study limitations. Finally, this chapter will end with suggestions for future research.

**Implications for practice.** One of the main advantages of the SRAT:MMA is group administration as a self-report survey as opposed to individual administration as it
is done in full microanalytic methods. Researchers can use this method to measure student self-regulation and teachers can administer this scale along with a meaningful task to diagnose group or individual student self-regulation deficiencies. As a class, professors and teachers can use this tool to evaluate their students’ self-regulation strengths and weaknesses and target those areas either through direct intervention, or modeling techniques (e.g., stressing the importance of comparing your outcomes to your own goals as opposed to your classmates, or emphasizing the steps to solving a problem as opposed to getting the answer right).

Even though this method was designed for assessing self-regulation as a class, individual students’ self-regulation deficiencies may also be identified. Specifically, if a professor or teacher notices that there is one student or a small group of students struggling in the classroom, then this measure can be administered to identify areas in which their self-regulation can be improved to increase performance. In terms of researchers, self-regulation can be assessed meaningfully and holistically by administering a relatively simple self-report survey. Although the SRAT:MMA, by nature, is a self-report survey, it deviates from classic self-report surveys in that it is highly contextualized and takes into account that self-regulation is both an event and skill.

In this present study, students were told one week ahead that there would be an extra credit opportunity during reading week. As a result, some students may have engaged in different strategies to study for the test prior to the testing date. However, because student engagement in this process (or lack thereof) occurred outside of the
assessment window, task strategies were not measured. Researchers should take this into account when using a modified microanalytic approach. Specifically, the measures should include only the self-regulation processes that students actively engage in at the time the measure is administered. This may lead to different data collection phases, where student forethought phases are collected at one point in time and performance and self-reflection phase processes are collected at a different point in time.

There are important considerations to keep when using this scale. Due to the context and task specific nature of this method of measurement, this scale must be administered along with a specific task. As opposed to general survey measures that can be administered without an accompanying task, this method of measurement requires a meaningful task in which students can anchor their self-regulation responses to. Further, the task should also be designed to require students to engage in forethought, performance, and self-reflection processes. For example, if forethought phase measures were the only processes administered, then student self-regulation was not truly assessed—because self-regulation involves more than just motivational beliefs. This will require more effort on the researchers to collect data from students within a meaningful context and task.

**Research limitations.** It is important to note that the sample may have been biased. Students were given the option to attend a study session during reading week. The students who chose to attend the session and thus participated in the study may have been inherently more self-regulated than students who did not attend the session. There may have been more variation in the data if attendance was required, leading to more robust,
less potentially biased sample. Further, it was unknown whether the student participants were also enrolled in other math courses during the same semester. Students who were only enrolled in one math class may exhibit different self-regulation profiles than students who were enrolled in multiple math classes. Therefore, future research should control for various enrollment data when investigating the self-regulation patterns within a given domain.

Another limitation is that the non-consequential nature of the task may have compromised the accuracy of the self-regulation measures due to the decreased effort expended by the student to do well on the test (i.e., students are more likely to engage in higher levels of self-regulation when a task is consequential than when a task is not consequential [Sundre & Kitsantas, 2004]). As previously discussed, the non-consequential nature of the task may have also influenced the fidelity of the stereotype threat manipulation.

It is important to emphasize that the results regarding high and low self-regulation should be interpreted with caution. First, due to the high number of statistical tests performed on the same data (i.e., 18 different comparisons), the alpha level is artificially inflated, which leads to an increase in Type I error. Second, a third of the data set was removed when the continuous self-regulation variables were dichotomized. Dichotomization of a continuous variable is generally advised against, given the methodological and empirical issues that typically arise when doing so (i.e., reduced statistical power and increase in Type II error [Cohen et al., 2003]). However, the rationale for the number of significance tests and dichotomization was to investigate
whether achievement differences emerged at each extreme end of self-regulation. Doing so provided the initial evidence of validity regarding the ability of the SRAT:MMA to detect self-regulation deficiencies. Future research may want to further investigate the initial findings of this present study with regression and ANOVA procedures to address these empirical limitations.

There are also limitations with regards to how the achievement measure was used in this study. Specifically, achievement on the math test was used as: 1) an indicator of effective versus ineffective self-regulation practices, 2) a moderator of the relationship between different phases of self-regulation, 3) the dependent variable in the predictiveness of the self-regulation processes as measured by the SRAT:MMA, and 4) an indicator of stereotype threat performance differences. This was done because students were guaranteed anonymity and that no other achievement indicators would be collected. As a result, achievement on the practice math test was used as a proxy for prior and future performance, which may not be theoretically accurate given the context of this study. Therefore, future research should attempt to collect multiple indicators of performance to more accurately investigate the relationship between self-regulation and achievement. Finally, the last limitation is that generalizations of this study are limited to college students enrolled in a non-introductory math course.

Suggestions for future research. Although the results regarding the ability of SRAT:MMA to assess student self-regulation are promising, future research should further investigate the validity of the scales in different domains such as history or English as well as in introductory courses. Further, researchers should also investigate the
validity of this approach with different types of academic tasks. There is evidence in previous research (DiBenedetto & Zimmerman, 2010, 2013) that microanalytically measured self-regulation processes may explain achievement differences in abstract, conceptual academic tasks more so than in straightforward academic tasks (short answer or multiple choice tests based on recall). Therefore, it would be interesting to see whether similar patterns in this modified microanalytic method emerge, where self-regulation processes measured with the SRAT:MMA can better predict achievement in abstract tasks than in straightforward tests.

As previously mentioned, one of the limitations of this study was that student attendance was not required during experimentation. Therefore, it was unknown whether the students who chose to attend were inherently more self-regulated or not. Future research should address this issue by accessing student final course grades or final test grades. Doing so would allow researchers to: 1) investigate whether there were achievement differences between students who participated in the study and students who did not participate in the study, and 2) investigate whether achievement on the test during experimentation is representative of achievement on an actual, consequential math test.

Another topic for future research is to explore specific task value versus general task value. In this study, it was found that math domain task value was more predictive of test scores than the value placed on doing well on the test. According to educational researchers, the more context specific the self-regulation processes are, the more power it will have to predict student achievement (Cleary, 2011; Winne & Perry, 2000). In this study, however, the math general task value measure was actually more predictive of
student achievement than the task specific measure of task value. Although this may potentially be due to the context of the study (i.e., extra credit, non-consequential test administered during an optional study session), this topic may warrant further research.

Item development may also be another topic for future research. In this study, it was a priority to use items from previous microanalytic studies and validated survey instruments. This was done to increase the content validity of this scale, given that those previous microanalytic and global scale items had already been validated. However, it can be argued that the microanalytic methodology inherently allows for researchers to create items specifically for the purposes of the study (Cleary, 2011). Therefore, future research should investigate whether the psychometric properties differ according to prior microanalytic and validated survey items versus purely researcher created items following Cleary’s (2011) microanalytic item creation guidelines. For example, researchers could administer the SRAT:MMA with one sample of students, but administer a separate microanalytic survey using only researcher created items with another sample of students. Following, researchers could then investigate whether the created items provide more or equally as accurate self-regulation information than already validated items.

Overall, it is important to reiterate that the results in this study show only initial evidence of validity. This was not full scale validity study and further research with a larger sample size is required to confirm whether it is comparable to full microanalytic approaches or more beneficial than global self-report scales. This is especially important given that a third of the data was essentially dropped when the self-regulation variables
were dichotomized into high and low groups. Although the rationale to dichotomize self-regulation was to determine whether high and low levels of self-regulation practices would result in differing levels of achievement, future research should use other methods of analyses, perhaps with regression techniques, to investigate self-regulation differences between achievement levels. Additionally, in order to further confirm the validity of the SRAT:MMA, future researchers should collect self-regulation data using global self-report instruments, the SRAT:MMA, and classic microanalytic methods to investigate the differences in validity between the three methods of measurement.

Finally, in terms of stereotype threat, the fidelity analyses unfortunately showed that the manipulation was unsuccessful in invoking stereotype threat in this sample of students. However, just because the manipulation failed does not indicate that stereotypes do not play a role in student achievement or learning behaviors. Although prior research has confirmed that stereotype threat is both difficult to detect and confirm, it continues to negatively influence student achievement (Cromley et al., 2013; Good et al., 2008; Nguyen & Ryan, 2008; Stoet & Geary, 2012). This may lead to several philosophical and methodological questions such as: How does stereotype threat operate in naturalistic settings; if the stereotype threat effect is robust, why is it more commonly detected in highly controlled laboratory settings than naturalistic settings; and, maybe most importantly, if self-regulation is a powerful determinate of academic achievement that emphasizes the self-sufficient role of the learner to overcome barriers to learning, how can this concept be used to help students overcome the barrier of gender and racial stereotypes?
APPENDIX A

IRB Approval Letter

TO: Anastasia Kitsantas, Psychology
FROM: Aurali Dade
Assistant Vice President, Research Compliance

PROTOCOL NO.: 8498 Research Category: Doctoral Dissertation

PROPOSAL NO.: N/A

TITLE: Exploring the effect of Self Regulation During Test taking on College Women’s Experience of Stereotype Threat in Upper Level Mathematics

DATE: April 25, 2013

Cc: Faye Hule

Under George Mason University (GMU) procedures, the amendment submitted on April 24, 2013, to this protocol does not change the status of the project. This project remains exempt since it falls under DHHS Exempt Category 2, research involving the use of educational tests (cognitive, diagnostic, aptitude, achievement), survey procedures, interview procedures or observation of public behavior.

A copy of the revised approved consent form is attached. Please use this copy with the approval stamp in your research.

You may proceed with data collection. Please note that any further modifications to your protocol must be submitted to the Office of Research Integrity & Assurance (ORIA) for review and approval prior to implementation. Any adverse events or unanticipated problems involving risks to subjects including problems with confidentiality of data identifying the participants must be reported to the GMU Office of Research Integrity & Assurance.

GMU is bound by the ethical principles and guidelines for the protection of human subjects in research contained in The Belmont Report. Even though your data collection procedures are exempt from review by the GMU IRB, GMU expects you to conduct your research according to the professional standards in your discipline and the ethical guidelines mandated by federal regulations.

Thank you for cooperating with the University by submitting this protocol for review. Please call me at 703-993-5381 if you have any questions.
APPENDIX B

The Self-Regulation Assessment Tool: A Modified Microanalytical Approach

Part A: Academic Motivation

Directions: Before you take your math test, I would like you to reflect on your different motivations and goals. In the following questions, some are open-ended (where you write out your perceptions) and some are on a scale (for example, 1-7 response scale). In those scale questions, please circle the number that best represents how you feel. Please answer these questions in the order that it is given to you as accurately and honestly as possible. Remember, there are no right or wrong answers and your identity is anonymous.

1) What is the grade that you want to achieve on the math test you are about to take? (Out of five total questions): __/5

2) On a scale from 1 to 7, how difficult will it be for you to achieve this grade on this test?

<table>
<thead>
<tr>
<th>Not very Difficult</th>
<th>Very Difficult</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td></td>
</tr>
</tbody>
</table>

3) Is it more important for you to focus on: (circle a or b)
   a) correctly following the steps to solving the problem or
   b) getting the answer right

4) On a scale from 1 to 7, how confident are you that you will be able to reach the grade you want to achieve on this math test?

<table>
<thead>
<tr>
<th>Not very Confident</th>
<th>Very Confident</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td></td>
</tr>
</tbody>
</table>

5) Are you planning on using any strategies to help you succeed on this test? (For example, any test taking strategies, techniques, methods, or line of attacks?)

Yes  No

If yes, what strategies do you plan on using?

6) How important is it for you to do well on this math test?

<table>
<thead>
<tr>
<th>Not very Important</th>
<th>Very Important</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td></td>
</tr>
</tbody>
</table>

Continue on with your math test on the next two pages
<table>
<thead>
<tr>
<th></th>
<th>How confident are you about your understanding of question 1?</th>
<th>Not very Confident</th>
<th>Very Confident</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>How confident are you that you have answered question 1 correctly and completely?</td>
<td>Not very Confident</td>
<td>Very Confident</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>How confident are you about your understanding of question 2?</td>
<td>Not very Confident</td>
<td>Very Confident</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>How confident are you that you have answered question 2 correctly and completely?</td>
<td>Not very Confident</td>
<td>Very Confident</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>How confident are you about your understanding of question 3?</td>
<td>Not very Confident</td>
<td>Very Confident</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>How confident are you that you have answered question 3 correctly and completely?</td>
<td>Not very Confident</td>
<td>Very Confident</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>How confident are you about your understanding of question 4?</td>
<td>Not very Confident</td>
<td>Very Confident</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>How confident are you that you have answered question 4 correctly and completely?</td>
<td>Not very Confident</td>
<td>Very Confident</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>How confident are you about your understanding of question 5?</td>
<td>Not very Confident</td>
<td>Very Confident</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>How confident are you that you have answered question 5 correctly and completely?</td>
<td>Not very Confident</td>
<td>Very Confident</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
</tbody>
</table>

STOP HERE

Please flip over the entire survey packet and raise your hand once you are done with the test. One of us will come and grade it. Do not continue with the survey until further instructions.
Part B: Post-test reflections

Thank you for taking the test and responding to the different questions. Now, I would like you to reflect on your performance by answering these following questions. Please answer as honestly as possible and remember, there are no right or wrong answers and your identity is anonymous.

7) At any point during this test, did your concentration wander? **Yes** **No**
   If yes, what did you do to refocus your concentration?

Please report your grade on the math test: ____/5

8) Please complete the following sentence: "I feel that my performance on this test was successful/unsuccesful because (Circle a, b, or c below)."
   a) I did/didn’t do better than the class average
   b) I did/didn’t do better than my previous tests
   c) I did/didn’t meet my goal

9) Please describe why you think you received this particular grade on this math test.

10) How satisfied are you with your score on this math test? Not very Satisfied 1 2 3 4 Very Satisfied 5 6 7

11) Was this test biased? Not at all 1 2 3 4 5 6 7 Extremely 8 9 10 11 12 13 14 15

12) How difficult was this test? Not very difficult 1 2 3 4 Very difficult 5 6 7

13) How hard did you try to do well on this test? Not very hard 1 2 3 4 Very hard 5 6 7

14) Do you consider yourself good at math? No, not very good 1 2 3 4 Yes, very good 5 6 7

15) How important is it for you that you are good at math? Not very important 1 2 3 4 Very Important 5 6 7
Part C: Demographic Information
Gender:   Male       Female
Age: __________
What is your ethnic background (circle one):
   a) African American
      Asian:
      b) East Asian
      c) South East Asian
   d) Caucasian
   e) Hispanic
   f) Middle Eastern
   g) Indian
   h) Mixed (please specify): ______________
   i) Other (please specify): ____________
What is your current major? ________________
How many math courses have you completed at the college level? ________
What year are you?: (Circle one)   Freshman   Sophomore   Junior
                     Senior
Are you currently enrolled in any of these math courses: (Check all the ones you are enrolled in)
   □  MATH 213  Analytic Geometry & Calc III
   □  MATH 301  Number Theory
   □  MATH 325  Discrete Mathematics II
   □  MATH 322  Advanced Linear Algebra
   □  MATH 321  Abstract Algebra
   □  MATH 315  Advanced Calculus I
   □  MATH 314  Introduction to Applied Math
   □  MATH 316  Advanced Calculus II
   □  MATH 351  Probability
   □  MATH 352  Statistics
   □  None of the above
APPENDIX C

Professor Created Practice Tests

Class 1: Math 114

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

Use the shell method to find the volume of the solid generated by revolving the region bounded by the given curves and lines about the y-axis.

1) \( y = \frac{2}{\sqrt{x}} \), \( y = 0 \), \( x = 1 \), \( x = 4 \)

A) \( \frac{64}{3} \pi \)  
B) \( 24 \pi \)  
C) \( \frac{56}{3} \pi \)  
D) \( \frac{28}{3} \pi \)

Evaluate the integral.

2) \( \int 11x \sin x \, dx \)

A) \( 11 \sin x - 11x \cos x + C \)  
B) \( 11 \sin x + 11x \cos x + C \)  
C) \( 11 \sin x + 11x \cos x + C \)  
D) \( 11 \sin x - x \cos x + C \)

Express the integrand as a sum of partial fractions and evaluate the integral.

3) \( \int \frac{3x + 23}{x^2 + 6x + 5} \, dx \)

A) \( \ln \left| \frac{(x + 1)^3}{(x + 5)^2} \right| + C \)  
B) \( \ln \left| \frac{(x + 1)^2}{(x + 5)^3} \right| + C \)  
C) \( \ln \left| \frac{(x + 1)^5}{(x + 5)^2} \right| + C \)  
D) \( \ln \left| \frac{(x + 1)^3}{(x + 5)^5} \right| + C \)

Find the interval of convergence of the series.

4) \( \sum_{n=0}^{\infty} \frac{(x - 7)^n}{1 + 2n} \)

A) \( 6 \leq x < 8 \)  
B) \( 4 < x < 10 \)  
C) \( 5 \leq x \leq 9 \)  
D) \( 5 < x < 9 \)

Evaluate the improper integral or state that it is divergent.

5) \( \int_{-\infty}^{-2} \frac{5}{x^4} \, dx \)

A) Divergent  
B) \( \frac{5}{24} \)  
C) \( \frac{5}{96} \)  
D) \( -\frac{5}{8} \)
Class 2: Math 114

1. Evaluate improper integral

\[ \int \frac{dx}{x(x+1)^7} \]

a) \(-\infty\)
b) \(-\frac{1}{6} \)
c) \frac{1}{2} \)
d) \(-\frac{1}{6} \)

2. Indefinite integral \( \int e^x \sin(x) \, dx \) is equal to:
   a. \( e^x (\sin(x) + \cos(x)) + C \)
   b. \( e^x (\sin(x) - \cos(x)) + C \)
   c. \( \frac{1}{2} e^x (\sin(x) - \cos(x)) + C \)
   d. \( e^x (\cos(x) - \sin(x)) + C \)

3. Which of the following statements is true for the series

\[ \sum \frac{(-1)^k k}{2k^2 + 1} \]

a. Converges absolutely
b. Converges conditionally
c. Diverges

4. Interval of convergence for the power series

\[ \sum \frac{(x + 1)^k}{\sqrt{k}} \]

is given by:
   a. \((\infty, \infty)\)
   b. \((-1, 1)\)
   c. \((-2, 0)\)
   d. \((-2, 0)\)

5. Which of the following gives a parametric description of the line segment from point \( P(0,0) \) to point \( Q(2,3) \):
   a. \( x = t, y = 3t \)
   b. \( x = 1 + 2t, y = 1 + 3t \)
   c. \( x = 2t, y = 3t \)
   d. \( x = t, y = 2t - 1 \)
Class 3: Math 213

1. For the vector field \( F = < 1, (x + yz), (xy - \sqrt{z}) > \), the curl and divergence are:
   a. \( z - \frac{1}{2\sqrt{z}} \) and \( (x - y) i - yj + k \) respectively
   b. \( (x - y) i - yj + k \) and \( z - \frac{1}{2\sqrt{z}} \) respectively
   c. \( z - yz j + k \) and \( (x - y) i - yj + k \) respectively
   d. \( (x + xy) i - yj + k \) and \( z - \frac{1}{2\sqrt{z}} \) respectively

2. Use polar coordinates to find the volume of a sphere of radius \( a \)
   a. \( \frac{3}{64\pi} \)
   b. \( \frac{3}{(64\pi^2)} \)
   c. \( \frac{3}{4} \pi a^2 \)
   d. \( \frac{3}{4} \pi a^3 \)

3. The equation of the plane through the point \((2,4,-1)\) with normal vector \( n = <2,3,4> \) is:
   a. \( 2x+3y+4z=12 \)
   b. \( 2x+2y-3z=12 \)
   c. \( 2x-3y+4z=12 \)
   d. \( 2x+4y+4z=-12 \)

4. The limit of the function \( f(x,y) = \frac{xy \cos y}{3x^2+y^2} \) is:
   a. \( \frac{1}{3} \)
   b. 1
   c. 0
   d. Does not exist

5. The point \((1,-1,4)\) is given in rectangular coordinates. The same point expressed in cylindrical coordinates is:
   a. \( (2, \frac{4\pi}{3}, 2) \)
   b. \( (2, \frac{4\pi}{3}, 2) \)
   c. \( (\sqrt{2}, \frac{7\pi}{4}, 4) \)
   d. \( (\sqrt{2}, \frac{4\pi}{3}, 2) \)
Class 4: Math 352

1. A sufficient statistic for \( \theta \) in a random sample of size \( n \) from the pdf \( f(x; \theta) = \theta x^{\theta-1} \) \( 0 < x < 1 \) is:
   a. \( \prod \ln xi \)
   b. \( \sum xi \)
   c. \( \prod xi \)
   d. \( \ln \sum xi \)

2. For \( f(x; \theta) = \frac{1}{\theta} 0 < x < \theta \) the unbiased estimator of \( \theta \) is:
   a. \( X_{max} \)
   b. \( X_{min} \)
   c. \( \frac{n+1}{n} X_{max} \)
   d. \( \frac{n}{n-1} X_{max} \)

3. If \( H_0: \mu = 30 \) is tested against \( H_1: \mu > 30 \) using \( n = 16 \) observations (normally distributed) and \( 1 - \beta = 0.85 \) when \( \mu = 34, what does \( \alpha \) equal? Assume \( \sigma = 9 \)
   a. 0.05
   b. 0.23
   c. 0.31
   d. 0.01

4. For a particular bivariate data set \((X,Y)\) a least squares line was derived as: \( \hat{Y} = 6 - 0.5X \).
   Which of the following values could be the correlation coefficient between \( X \) and \( Y \)?
   a. 0.5
   b. -0.23
   c. 0.9
   d. 0

5. Records kept at a racetrack showed the following distribution of winners as a function of their starting post position (for a total of 100 races).

<table>
<thead>
<tr>
<th>Starting post</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of winners</td>
<td>35</td>
<td>15</td>
<td>10</td>
<td>28</td>
<td>12</td>
</tr>
</tbody>
</table>

If we are using a goodness of fit test to test the hypothesis that all posts are equally likely, what would be the expected number of winners for post 4?
   a. 8
   b. 25
   c. 10
   d. 20
1. Suppose that $X$ and $Y$ are independent uniform random variables on the interval $[0, 1]$. Find $P(X + Y) \leq \frac{1}{4}$.

2. Suppose that $X$ is uniformly distributed on $[0, 1]$. Find the moment generating function of $X$.

3. Suppose that an urn contains one red ball and one green ball. One ball is drawn from the urn. Let $X$ be 1 if the ball drawn is red, 0 if it is green. Let $Y$ be 1 if the ball drawn is green, 0 if it is red. Find $Cov(X, Y)$.

4. Suppose that $X_1, X_2, \ldots, X_{100}$ are independent, identically distributed random variables, and that $E[X_1] = Var(X_1) = 1$. Find $E[M_{100}]$ and $Var(M_{100})$, if $M_{100} = \frac{X_1 + X_2 + \cdots + X_{100}}{100}$.

5. Let $X$ be a random variable with $E[X] = Var(X) = 1$. What does Chebyshev say about $P(|X - 1| \geq 3)$?
APPENDIX D

Informed Consent

The Mental Processes Underlying Mathematics Abilities in Students Enrolled in Higher Level College Mathematics Courses

RESEARCH PROCEDURES
This research is being conducted to investigate the mental processes underlying math ability. If you agree to participate, you will be asked to complete a series of surveys aimed at assessing your motivation and learning behaviors in the class. This survey will be given to you in conjunction with your mathematics practice test. Your total participation would last about 25 - 40 minutes.

RISKS
There are no foreseeable risks for participating in this research.

BENEFITS
There are no benefits to you as a participant other than to further research in mathematics and learning.

CONFIDENTIALITY
The data in this study will be anonymous. You will be assigned an ID number, which will not be linked to your identity.

PARTICIPATION
Your participation is voluntary, and you may withdraw from the study at any time and for any reason. If you choose to participate, you will be asked to complete a survey which will take approximately 5-10 minutes to complete. The math test will take approximately 20-30 minutes to complete, depending on the time allotted by your professor. If you decide not to participate or if you withdraw from the study, there is no penalty or loss of benefits to which you are otherwise entitled. If you choose not to participate, you will only be given the math test. There are no costs to you or any other party.

ALTERNATIVES TO PARTICIPATION
If you choose not to participate, you will be given only the math test to complete.

CONTACT
This research is being conducted by Faye Huie and Dr. Anastasia Kitsantas at George Mason University. They may be reached at (703) 989-3294, or (703) 993-2688 for questions or to report a research-related problem. You may contact the George Mason University Office of Research Integrity & Assurance at 703-993-4121 if you have questions or comments regarding your rights as a participant in the research.

This research has been reviewed according to George Mason University procedures governing your participation in this research.

APPROVED

George Mason University
APPENDIX E

Student Debriefing Form

Debriefing Form

Exploring the Effect of Self-Regulation during Test Taking on College Women’s Experience of Stereotype Threat in Upper Level Mathematics

Dear Participant,

During this study, you were asked to take a mathematics test as well as respond to a series of questionnaires regarding your motivation and self-regulation. You were told that the purpose of the study was to investigate the mental processes underlying mathematics ability. The actual purpose of the study was to examine stereotype threat effects for women in mathematics, and how it may be eliminated through high levels of self-regulated learning. You will be given extra credit points for your participation. The number of extra credit points will be determined by your professor.

We did not tell you that stereotype threat was the purpose of the study because it would have compromised the stereotype threat effect and invalidate the entire purpose of this study.

You are reminded that your original consent document included the following information:
- Your participation is voluntary;
- Your identity and data is completely anonymous; and
- You are entitled to any benefits to which you are otherwise entitled if you withdraw from the study.

If you have any concerns about your participation or the data you provided in light of this disclosure, please discuss this with me. I will be happy to provide any information I can to help answer questions you have about this study.

If your concerns are such that you would now like to have your data withdrawn, I will do so.

If you have questions about your participation in the study, please contact me at fhuie@gmu.edu (703-989-3294), or my faculty advisor, Dr. Anastasia Kitsantas at akitsantas@gmu.edu (703-993-2688).

If you have questions about your rights as a research participant, you may contact the George Mason University Office of Research Integrity & Assurance at 703-993-4121.

Please again accept our appreciation for your participation in this study.

APPROVED

George Mason University

160
## APPENDIX F

### Overall Self-Regulation by Stereotype Threat Condition Descriptive Statistics

<table>
<thead>
<tr>
<th>Self-Regulation Variables</th>
<th>Overall Sample</th>
<th></th>
<th>High Stereotype Threat</th>
<th></th>
<th>Low Stereotype Threat</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>N = 191</td>
<td></td>
<td></td>
<td>N = 96</td>
<td></td>
<td>N = 95</td>
<td></td>
</tr>
<tr>
<td>Achievement</td>
<td>2.74</td>
<td>1.39</td>
<td>2.86</td>
<td>1.43</td>
<td>2.62</td>
<td>1.35</td>
</tr>
<tr>
<td>Grade Goal</td>
<td>4.60</td>
<td>.70</td>
<td>4.60</td>
<td>.70</td>
<td>4.60</td>
<td>.71</td>
</tr>
<tr>
<td>Goal Difficulty</td>
<td>4.24</td>
<td>1.42</td>
<td>4.20</td>
<td>1.37</td>
<td>4.28</td>
<td>1.48</td>
</tr>
<tr>
<td>Process vs Outcome Goal</td>
<td>.71</td>
<td>.46</td>
<td>.72</td>
<td>.45</td>
<td>.70</td>
<td>.46</td>
</tr>
<tr>
<td>Self-Efficacy</td>
<td>4.21</td>
<td>1.40</td>
<td>4.46</td>
<td>1.26</td>
<td>3.95</td>
<td>1.48</td>
</tr>
<tr>
<td>Strategic Planning</td>
<td>.63</td>
<td>.92</td>
<td>.67</td>
<td>.97</td>
<td>.60</td>
<td>.88</td>
</tr>
<tr>
<td>Task Value: Math Test</td>
<td>4.79</td>
<td>1.88</td>
<td>4.78</td>
<td>1.85</td>
<td>4.80</td>
<td>1.93</td>
</tr>
<tr>
<td>Task Value: Math Domain</td>
<td>5.49</td>
<td>1.52</td>
<td>5.54</td>
<td>1.58</td>
<td>5.45</td>
<td>1.47</td>
</tr>
<tr>
<td>Metacognition: Understanding of Question</td>
<td>4.55</td>
<td>1.32</td>
<td>4.64</td>
<td>1.38</td>
<td>4.46</td>
<td>1.26</td>
</tr>
<tr>
<td>Metacognition: Confidence in Answer</td>
<td>4.08</td>
<td>1.30</td>
<td>4.11</td>
<td>1.31</td>
<td>4.04</td>
<td>1.29</td>
</tr>
<tr>
<td>Attentional Control</td>
<td>.39</td>
<td>.60</td>
<td>.34</td>
<td>.52</td>
<td>.44</td>
<td>.66</td>
</tr>
<tr>
<td>Self-Evaluation</td>
<td>2.48</td>
<td>.79</td>
<td>2.56</td>
<td>.74</td>
<td>2.41</td>
<td>.84</td>
</tr>
<tr>
<td>Attributions: Stability</td>
<td>.10</td>
<td>.30</td>
<td>.12</td>
<td>.33</td>
<td>.08</td>
<td>.28</td>
</tr>
<tr>
<td>Attributions: Controllability</td>
<td>.80</td>
<td>.40</td>
<td>.75</td>
<td>.44</td>
<td>.86</td>
<td>.35</td>
</tr>
<tr>
<td>Attributions: Internal vs External</td>
<td>.86</td>
<td>.35</td>
<td>.82</td>
<td>.39</td>
<td>.90</td>
<td>.30</td>
</tr>
<tr>
<td>Self-Satisfaction</td>
<td>3.38</td>
<td>2.14</td>
<td>3.51</td>
<td>2.23</td>
<td>3.25</td>
<td>2.05</td>
</tr>
</tbody>
</table>


170


Biography

Faye C. Huie graduated from Robinson High School, Fairfax, Virginia, in 2002. She received her Bachelor of Arts from George Mason University in 2006. She was employed as a substitute teacher in Fairfax County for two years and received her Master of Education from George Mason University in 2008.