On Mathematical Expertise, Inhibitory Control, and Facets of College Students' Psychoeducational Profile: An Empirical Investigation

Brian Darrow, Jr.

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2023

© 2023

Brian Darrow, Jr.

All Rights Reserved

Abstract

On Mathematical Expertise, Inhibitory Control, and Facets of College Students' Psychoeducational Profile: An Empirical Investigation

Brian Darrow, Jr.

Although the importance of problem solving as an essential component of mathematics learning and doing has consistently been recognized, recent research has only just begun to identify and describe the complex set of variables influencing the endeavor. Therefore, the aim of this study was to empirically investigate the relationships between several of these variables: mathematical expertise (as measured by the advanced nature of the mathematics courses students have taken, and are enrolled in), the cognitive ability known as inhibitory control (the ability to inhibit or suppress an immediate response to a stimulus, and engage in deeper, more reflective thought), and facets of college students' psychoeducational profile (e.g., academic habits of mind, future orientation, self-limiting beliefs), which provide information about the nature of college students' learning and development.

In this study, one hundred and thirty college students, enrolled in different levels of mathematics courses (from introductory courses to major courses in mathematics) were administered a modified version of the Cognitive Reflection Test (an instrument designed to measure the ability to activate one's inhibitory control capacities) and a survey instrument designed to measure domain-general and mathematics-specific psychoeducational facets of their academic profile. Information about membership to other subgroups (e.g., gender, academic major, mathematics courses taken in high school) helped to further contextualize the findings.

The majority of all participants did not correctly solve any of the problems of the modified version of the Cognitive Reflection Test which required inhibitory control. However,

those with a greater level of mathematical expertise (i.e., those taking more advanced mathematical courses) performed significantly better than their peers on these problems and exhibited more desirable responses on the psychoeducational survey instrument. Responses to items of the survey instrument that measured behaviors, habits, and experiences that limit students in their conception of, approach to, and engagement with mathematics indicate the presence of a psychoeducational facet specific to mathematics that cannot be sufficiently explained by domain-general facets also under measure. These limiting characteristics related to mathematics were also significantly related to students’ performance on the modified version of the Cognitive Reflection Test, indicating a potential relationship between such characteristics and problem solving success on inhibitory control tasks. Considering the measures of mathematical expertise utilized in the current study, the social nature of mathematics learning may help explain the development of both inhibitory control ability and limiting beliefs in mathematics.

The current study extended the methods utilized in previous research to examine the relationships between inhibitory control and mathematical expertise in college students while also investigating these in relation to particular psychoeducational variables known to influence learning and development of college students. The findings of this small-scale empirical study provide a modest step forward in these areas of research by providing another lens through which to view several phenomena already being extensively investigated by other researchers.

Table of Contents

List of Tables and Figures

Chapter 1: Introduction

1.1 Need for the Study

The endeavor of mathematical problem solving is at the core of the construction, dissemination, and application of mathematical knowledge. The National Council of Teachers of Mathematics or NCTM (the foremost society of school mathematics educators in the United States) states in their 2000 *Principles and Standards for School Mathematics* (which continues to drive educational reform initiatives over twenty years after is original publication) that “Problem solving is an integral part of all mathematics learning” (p. 52). Although the importance of problem solving as an essential component of mathematics learning has consistently been recognized by mathematics educators and reformers alike (e.g., NCTM, 2022; NGA, 2010), recent research has only just begun to identify and describe the complex set of variables influencing the endeavor.

In 1985, Alan Schoenfeld provided one of the first robust explanatory frameworks for mathematical problem solving that highlighted the influence of sociological and psychological variables. Since, research on the latter has revealed that cognitive processes such as metacognition (one’s own regulation of cognitive processes), self-regulation (active control over emotion, motivation, and cognition), and executive functioning (processes responsible for working memory, focus, action, and effort) are among the most important to problem solving (Cragg et al., 2017; Cragg & Gilmore, 2014; Desoete & De Craene, 2019; Schoenfeld, 1985, 2013, 2016).

Among the cognitive variables influencing problem solving are intuition and inhibition, which have been given a great deal of recent attention. Building upon major works on the study of intuition in science and mathematics (perhaps most notably Fischbein, 1987), more recent work suggests that cognition consists of two processes, namely Type 1 and Type 2 (Evans & Stanovich, 2013; Leron & Hazzan, 2006, 2009). In citing Evans and Stanovich (2013), Van Dooren & Inglis (2015) provide a succinct description of this:

> On Evans and Stanovich's account, a process can be classified as Type 1 if it does not require working memory resources, and as Type 2 if it does. Typically, Type 1 processes are fast, automatic, operate in parallel, and are independent of cognitive ability. In contrast Type 2 processes are slow, limited by working memory capacity, effortful, conscious, rule-based and are related to cognitive ability. (p. 717)

The immediacy of Type 1 processes and their independence of reflective thought can lead to errors in mathematical problem solving situations (Frederick, 2005; Leron & Hazzan, 2006; Lubin et al., 2013; Van Dooren & Inglis, 2015; Verschaffel et al., 2020). The ability to inhibit these initial, immediate, and intuitive responses is often referred to as inhibitory control or IC (Van Dooren & Inglis, 2015), which has been shown to influence the success of individuals of varying ages and differing levels of mathematical expertise on a wide range of mathematical tasks (Coulanges et al., 2021; Frederick, 2005; Lubin et al., 2013, 2016; Obersteiner et al., 2013, 2016; Rossi et al., 2019; Van Dooren & Inglis, 2015).

In recent decades, a number of instruments have been developed to measure inhibitory control; however, Frederick's (2005) Cognitive Reflection Test (CRT) seems to stand out in terms of its popularity and utility in measuring this in adults. The test consists of three mathematical word problems that are "hard to solve, not because of their linguistic or mathematical demands but rather because the correct answer requires the inhibition of a very easy, seemingly correct answer" (Verschaffel et al., 2020, p. 9).

Partly due to the nature of the CRT as an instrument consisting of mathematical word problems, recent research has investigated the influence of mathematical knowledge on individuals' performance on the instrument. However, these investigations have relied heavily on general measures of intelligence (e.g., the SAT/ACT, the Wonderlic Personnel Test, the Differential Aptitudes Test, the Need for Cognition Scale, Raven's Progressive Matrices, Wechsler scales) and numeracy (e.g., the numeracy scales of Lipkus et al., 2001; Schwartz et al., 1997; Fagerlin et al., 2007), which provide an inherently narrow and limited view of mathematical expertise. Additionally, the findings of previous research that indicate the CRT is more difficult than previously used numeracy scales (Weller et al., 2012) and measures a distinct cognitive construct separate from numeracy (Liberali et al., 2011) suggest the need for investigations of different measures of mathematical expertise in relation to the CRT.

Such measures have been utilized in the investigation of other instruments measuring inhibitory control. For example, Lubin et al., (2016) investigated inhibitory control task performance (arithmetic word problems) in groups of undergraduate "experts" in mathematics (students in a mathematics major program of study) and undergraduate non-experts in mathematics (students in humanities majors). In other studies, Obersteiner et al. (2013, 2016) investigated the performance of adult mathematicians (individuals with graduate degrees in mathematics) on inhibitory control tasks (natural number bias tasks) in comparison to non-mathematicians and school-aged students. However, we are not aware of any study that incorporates such measures of mathematical expertise in college students in an investigation of CRT performance, although one study conducted by Gómez-Chacón et al. (2014) studied high school students' CRT performance in relation to mathematics course performance and belief systems. In addition, we are not aware of any studies that investigate CRT performance in relation to a further granulation of mathematical expertise in college students to that just described—that is, one that does not focus solely on the stark difference between mathematics experts and non-experts. Therefore, to further current research, a study that investigates the influence of mathematical expertise—measured by the extent of collegiate mathematics learning and not by numeracy scales—on CRT performance is needed.

Research is also needed that investigates the relationship between college students' mathematical expertise and facets of their psychoeducational profile. Longitudinal cohort research on the latter has demonstrated the inextricable link between college students' learning and development (Ben-Avie & Darrow, 2019). This research has shown that the facets of college students' academic profiles that are amenable to change are more meaningful and consistent predictors of collegiate academic outcomes than those that cannot be influenced after college enrollment. Although such components of a students' psychoeducational profile as academic habits of mind, future orientation, academic self-concept, and self-limiting beliefs have been traditionally difficult to measure, "these factors are directly related to teaching and learning in and out of the classroom" (Ben-Avie & Darrow, 2019, p. 47) in a way that standardized measures of crystallized learning are not.

However, the domain-general psychoeducational measures described above are not specific enough to measure similar factors related to mathematics. Although a recent pilot study made a first attempt in developing an instrument to measure these factors in college students (Darrow, 2020), further research is needed to develop an instrument capable of measuring learning and developmental experiences that contribute to students' belief systems in mathematics. Moreover, what is needed is an instrument that not only accounts for mathematics specific forms of psychoeducational facets known from previous research to influence college student learning and development (e.g., academic habits of mind, self-limiting beliefs), but also one that points to attributes that are amenable to change. Such an instrument, if administered to students of varying backgrounds in mathematics and levels of expertise, would help describe individual differences among these students, thereby informing ways to improve learning and developmental experiences for students in mathematics.

Finally, these measures have also never been investigated in relation to inhibitory control, specifically as measured by CRT performance. Previous research on high school students has linked CRT performance and mathematics belief systems (Gómez-Chacón et al., 2014), however the measures of belief systems used by Gómez-Chacón et al. (2014) were distinctly different from the measures piloted in the study by Darrow (2020), which were built on the psychoeducational variables identified by Ben-Avie and Darrow (2019). In other words, it currently unknown whether inhibitory control is related to the psychoeducational variables identified by Ben-Avie & Darrow (2019) that consistently explain academic success in college. Moreover, it is currently unknown whether mathematics-specific forms of these characteristics are related to inhibitory control, and whether the nature of this relationship changes for individuals of differing backgrounds in mathematics or levels of mathematical expertise. Furthermore, a study of this kind would contribute to the field of mathematics education new research on college students' learning, development, and problem solving in mathematics.

Therefore, in summary, an empirical investigation of the potential links between inhibitory control as measured by CRT performance, mathematical expertise, and domain-general and mathematics-specific components of college students' psychoeducational profile would meaningfully contribute to the field of mathematics education new research on variables related to learning, development, and problem solving in mathematics.

1.2 Purpose of the Study

The purpose of this study was to explore the potential links between cognitive, social, and psychological variables that are known to be related to collegiate academic outcomes through the administration of a modified version of the Cognitive Reflection Test and a psychoeducational survey to college students of differing levels of mathematical expertise. Specifically, this mixed-methods empirical study sought to answer the following research questions:

1. How do students from different collegiate mathematics courses perform on the Cognitive Reflection Test (modified by the investigator to include two additional problems)?
2. How do students from different collegiate mathematics courses respond to a psychoeducational survey measuring domain-general academic habits of mind, future orientation, and academic self-concept, and mathematics-specific academic habits of mind, beliefs, and past academic experiences?
3. What is the relationship between the performance of students from different collegiate mathematics courses on the modified version of the Cognitive Reflection Test and their responses to a psychoeducational survey measuring domain-general academic habits of mind, future orientation, and academic self-concept, and mathematics-specific academic habits of mind, beliefs, and past academic experiences?

1.3 Procedure

1.3.1 Setting and Participants

The participants in this study were all adult (at least eighteen years of age or older), undergraduate students at the same diverse, four-year public university in the northeastern United States. Aligning with the Integrated Postsecondary Education Data System (IPEDS) definitions of race and sex, the institution under study is composed of approximately half white and half non-white students, with nearly two-thirds of the student population being female. Since the study aimed to investigate the potential relationship between college students' extent of mathematical knowledge, problem solving, and students' psychoeducational make-up, each participant was actively enrolled in an undergraduate mathematics class at the university under measure.

Furthermore, participants were selected according to the level of mathematics classes in which they were enrolled. Each mathematics course at the university under study requires the satisfaction of a prerequisite requirement. These generally take the form of certain score attained on a placement exam; the transfer of acceptable credits from another postsecondary or secondary institution; or the successful completion (subject to minimum grade requirements) of a prerequisite course. Therefore, a direct measure of mathematical expertise is required to advance from one course level to another. This yields a leveling structure where courses in successive levels are separated by measurable milestones of mathematical proficiency and expertise. The university defines the academic levels in mathematics in this way according to their academic program structure.

The subgroups were formed after consulting university faculty and the university's course catalog to determine each course's prerequisite requirements; rigor and depth of the mathematical content; and whether these satisfy different subsequent quantitative major requirements defined by the university's academic program structure. The subgroups of participants were students from introductory (commonly referred to as "remedial" or "developmental") courses in mathematics, which do not satisfy the university requirements in mathematics and serve as prerequisite courses for courses in the next level (e.g., introductory/intermediate algebra); mathematics courses that satisfy the general education requirement for students not in the STEM (science, technology, engineering, or mathematics) majors (e.g., elementary statistics, mathematics for elementary education, etc.); mathematics courses required of STEM majors (e.g., calculus I, calculus II, intermediate statistics, etc.); and courses for majors in mathematics (e.g., foundations of mathematics, discrete mathematics, real analysis, etc.). Accordingly, in the order just described, the mathematics course level subgroups have been named, and will be referred to as the developmental level, the general level, the STEM level, and the mathematics level.

The resulting sample was selected by convenience. Willing faculty members in the mathematics department of the university created opportunities for the research to in their regularly-scheduled classrooms. Then, students from these classes were given the opportunity to voluntarily participate in the research. Data collection occurred between the dates of 4/20/2022 and 5/2/2022, which spanned the thirteenth, fourteenth, and fifteenth weeks of the university's semester. Therefore, students who were not enrolled in mathematics courses at this time were not eligible to participate. The withdrawal deadline for the university under study occurs during the tenth week of classes, and thus, students who withdrew from the class were not eligible to participate. Also, since the mathematical and psychoeducational instruments were intended to be given on paper, students who were enrolled in remote (virtual synchronous and asynchronous) classes were not invited to participate.

In total, 130 students elected to participate in the study. Of these 130, 49 (37.69%) were male, 78 (60%) were female, 1 (0.77%) preferred not to provide gender information, and 1 (0.77%) identified as non-binary. Gender information was recorded since several studies in previous literature note that gender is an important variable in the analysis of CRT scores (e.g., Brañas-Garza et al., 2019; Campitelli & Gerrans, 2014; Frederick, 2005). In terms of academic year, 46 (35.38%) students identified as being in their first-year, 26 (20%) were in their sophomore year, 34 (26.15%) were in their junior year, 22 (16.92%) were in their senior year, and 1 (.77%) identified as not belonging to a particular academic year.

In terms of the course leveling structure, students from each course level (developmental, general, STEM, and mathematics) were recruited. In total, 11 (8.46%) were from the developmental level, 55 (42.31%) were from the general level, 44 (33.85%) were from the STEM level, and 20 (15.38%) were from the mathematics level. The courses from which these students included introductory/intermediate algebra (developmental level); elementary statistics and mathematics for elementary education (general); calculus I, calculus II, and intermediate statistics (STEM); and foundations of mathematics, discrete mathematics, and real analysis (mathematics).

1.3.2 Data Collection

As mentioned above, data were collected in the classrooms of students' regularly-scheduled courses at the university under study during the thirteenth, fourteenth, and fifteenth weeks of the Spring 2022 semester during between the dates of 4/20/2022 and 5/2/2022. IRB approval for the study was sought and obtained from both Teachers College, Columbia University and the institution under measure. The participants were given a packet of paper that contained mathematical tasks and a psychological survey. The participants were informed that they would be asked to solve five mathematics problems varying in difficulty before responding to the survey. The participants were informed that they may use pen or pencil for both parts, they may use any free space in the packet for their mathematical work (free space was left underneath each problem), that they may use a calculator, that they may take any or all of the twenty-five minutes allotted for both tasks, and that they may stop at any time. All 130 participants responded to each of the mathematics problems and to most items on the psychoeducational survey, all finished their work within twenty minutes, and none elected to stop participating.

Each packet contained a modified version of the Cognitive Reflection Test (MCRT; given in Appendix A of this document) and the Short Mathematics Psychoeducational Inventory (SMPI; given in Appendix B of this document). The former consisted of five mathematics word problems with spaces provided for participants to engage in scratch work, to provide their final answers, and to indicate whether they have seen the problem before. The latter consisted of a 41-item survey questionnaire comprises demographic and subgroup items as well as 5-point Likert scale items measuring level of agreement to a set of statements.

1.4 Method

1.4.1 Research Question 1

To investigate the first research question, a modified version of the Cognitive Reflection Test (henceforth referred to as the MCRT, which is included in Appendix A of this document) was administered on paper to participating students. The MCRT consisted of the three original CRT problems and two decoy problems. To be successful on the CRT problems, one must activate their inhibitory control capacities and engage in cognitive reflection to effectively inhibit the intuitive, yet faulty response suggested by the problem structure. After successful inhibition of the intuitive response, the solver must then activate their more reflective, Type 2 cognitive processes and perform elementary mathematics operations to determine the correct response. Therefore, performance on these problems (including correct responses and the analysis of incorrect responses) provides a measure of the participants' ability to effectively activate inhibitory control capacities and engage in Type 2 thinking to correctly solve the problems.

Descriptive and inferential statistics were used to analyze the nature of participants' performance on the MCRT and determine differences across subgroups. When the statistical assumptions for parametric analyses were met, procedures such as Chi Square Tests for Independence were used. However, the small sample size of several subgroups and the non-normal nature of MCRT data required the use of non-parametric analyses such as mean square contingency, Spearman, and Kendall's Tau analyses for association; Mann-Whitney U-Tests (also known as the Mann-Whitney-Wilcoxon Test, or the Wilcoxon rank-sum test) for differences between two subgroups; and Kruskal-Wallis One Way Analyses of Variance (also known as the Kruskal-Wallis Test, or the Kruskal-Wallis H-Test) for differences across several subgroups.

1.4.2 Research Question 2

As was previously mentioned, after completing the MCRT, a comprehensive psychoeducational paper survey instrument was administered to the participants. All psychoeducational items (5-point Likert scale: strongly disagree, disagree, neutral, agree, strongly agree) appearing on the current survey have been previously tested through prior administration (Ben-Avie et al., 2012; Ben-Avie & Darrow, 2019; Darrow, 2016, 2020). These items were developed to measure domain-general and mathematics-specific psychoeducational factors such as academic habits of mind, future orientation, academic self-concept, and self-limiting beliefs. General demographic information was also collected for the purposes of identifying and analyzing subgroup membership. Henceforth, the survey described here will be referred to as the Short Mathematics Psychoeducational Inventory (SMPI), which is included in its entirety in Appendix B of this document. The latent structure of the SMPI was first analyzed using an exploratory factor analysis. In particular, a principal axis factoring procedure with a Promax rotation was utilized due to the non-normal data of the SMPI and since any factors emerging from the analysis were anticipated to be correlated to one another. Reliable psychometric scales were developed from the results of the factor analysis, which were then used to create composite scores.

Additionally, descriptive and inferential statistics were used to analyze the nature of participants' responses and determine differences across subgroups. Due to the ranked, non-normal nature of the data, including composite scores, non-parametric analyses were used, including Spearman, and Kendall's Tau analyses for association; the Mann-Whitney U-Tests (also known as the Mann-Whitney-Wilcoxon Test, or the Wilcoxon rank-sum test) for differences between two subgroups; and Kruskal-Wallis One Way Analysis of Variance (also known as the Kruskal-Wallis Test, or the Kruskal-Wallis H-Test) for differences across several subgroups.

1.4.3 Research Question 3

Participants' performance on the MCRT and their responses on the psychoeducational survey (SMPI) were compared across all subgroups. In essence, differences in mathematical expertise, the ability to activate inhibitory control capacities in mathematical problem solving situations, and elements of participants' psychoeducational profile were compared to determine the nature of any relationships that exist among these. Since the data of both the MCRT and the SMPI are ranked in nature, descriptive statistics and the non-parametric procedures previously described were leveraged here as well. Additionally, logistic regression analyses were conducted to predict zero and non-zero scores on the CRT subset of the MCRT using mathematics course level subgroup membership and composite scores of the SMPI.

Chapter 2: Review of Literature

2.1 Introduction

The endeavor of mathematical problem solving is at the core of the construction, dissemination, and application of mathematical knowledge. Bruner (1966) notes that "knowledge is a process, not a product" (p. 72), and in the case of mathematics, that process is problem solving. From an educational perspective, problem solving continues to be a focal point of the teaching and learning of mathematics. The National Council of Teachers of Mathematics or NCTM (the foremost society of school mathematics educators in the United States) states in their 2000 *Principles and Standards for School Mathematics* (which continues to drive educational reform initiatives over twenty years after is original publication) that "Problem solving is an integral part of all mathematics learning" (p. 52).

In their modern executive summary of the aforementioned document, NCTM notes that "by solving mathematical problems students acquire ways of thinking, habits of persistence and curiosity, and confidence in unfamiliar situations that serve them well outside the mathematics classroom" (NCTM, 2022, p. 4). The Common Core State Standards for Mathematics note that "mathematically proficient students can apply the mathematics they know to solve problems arising in everyday life, society, and the workplace" (NGA, 2010, p. 7). Such a conception of mathematical problem solving supports the widely-held position that mathematical knowledge can be transferred to, or utilized in inherently non-academic situations as a "practical skill" (Polya, 1945, p. 4). However, despite the prevailing notion that the "development of students as problem solvers is one of the most important functions of school" (Smith, 1973, p. 1), which continues to this day, research into understanding the complex phenomenon of mathematical problem solving is still, in many respects, in its infancy. The field of mathematics education has only just begun to understand the complex set of psychological, cognitive, and social variables influencing success in mathematical problem solving situations.

In this section, an introduction to the foundational research in mathematical problem solving will be given, followed by research on the role of self-regulation, metacognition, and executive functioning in mathematics learning and problem solving. The discussion then focuses on one particular cognitive construct and executive function, inhibitory control, which is the ability to inhibit immediate, intuitive, or default responses to a stimulus and engage in reflective, conscious, and higher-order thought. Research on inhibitory control and its emergence in mathematics education literature is detailed from the perspective of leading theories of higher cognition. The importance of this ability in mathematical learning and problem solving is highlighted, as are the advancements in its measurement.

The most popular and widely-used instrument for measuring inhibitory control (or cognitive reflection) and its use in studies linking it to a wide range of psychological and educational outcomes is reviewed. Then, the focus is shifted to recent research on college students' learning and development. Results from longitudinal cohort studies of the psychoeducational facets of college students' profile are detailed alongside discussions of the instruments utilized in this research. Finally, extensions of such research and instruments to the field of mathematics education and avenues of future exploration in this area are discussed.

2.2 Mathematical Problem Solving

As Jeremy Kilpatrick (2014a) notes, "Although mathematics has been taught and learned for millennia, not until the past century or so have the nature and quality of teaching and learning mathematics been studied in any serious manner" (p. 267). Normal schools, or schools dedicated to the training of teachers, in the United States have existed since the early 1800's; however, the professionalization of mathematics education and its formalization as an academic discipline did not occur until the early 1900's (Kilpatrick, 2014b). In the years following the conferral of the first doctoral degrees in mathematics education at Teachers College, Columbia University in 1906, research in the field steadily increased (Kilpatrick, 2014a).

Research in mathematics education is complemented by, and often comprises research from the social sciences. Research in psychology, in particular, has formed much of the foundation for mathematics education research dealing with the cognitive aspect of mathematics teaching, learning, and doing. However, much like the field of mathematics education itself, much of the research in the psychology of learning did not emerge until after the turn of the twentieth century. Dominowski & Bourne (1994) write that although "learning, memory, motivation, problem solving, and the like are basic processes in human behavior, these were simply not recognized as important topics of investigation before 1900" (p. 13). The years after 1900 saw major developments in the psychological research on thinking and problem solving (Dominowski & Bourne, 1994), which laid the foundation for mathematics education research investigating problem solving within the domain of mathematics teaching and learning.

Writings on mathematical problem solving have existed since antiquity; however, the formal research into the complex phenomenon has largely been formalized in the last sixty years (Kaur et al., 2009). Even as of the late 1960's, Kilpatrick (1969) noted that "problem solving is not now being investigated systematically by mathematics educators" (p. 523) and only a few hundred formal research reports had been published on the subject by that time. Although several influential works on the subject were published before this time (e.g., Binet, 1903; Blatt & Stein, 1959; Bloom & Broder, 1950; Buswell, 1956; Davis, 1966; Duncker, 1945; Gagné, 1965, 1966; Hadamard, 1954; Henderson & Pingry, 1953; Johnson, 1944; John, 1957; Kleinmuntz, 1966;

Lazerte, 1933; Newell et al., 1958; Rimoldi et al., 1964; Wertheimer, 1945; Wilson, 1967; Young, 1907), George Polya's 1945 text *How To Solve It* is widely considered to be among the most popular and influential in generating interest in such research.

One of Polya's goals was to not solely focus on the solutions of mathematical problems or finished results, but rather on the process that led to such results. The text provided lengthy discussions of problem situations and the processes that the solver, and the teacher of problem solving, often and ought to engage with for more desirable outcomes. The largest and presumably most popular takeaways from the text were Polya's ordered four-step problem solving process (understanding the problem; devising a plan; carrying out the plan; and looking back) and his discussion of "heuristics" or general problem solving techniques that could be employed in the solving of any problem. Alongside the interest generated by Polya's original (1945) and subsequent texts (Polya, 1954, 1965, 1974) in problem solving research, more attention was paid to what is formally meant by problem solving.

Building upon contributions from early researchers (e.g., Duncker, 1945, Henderson & Pingry, 1953), the modern conception of what constitutes a mathematical problem relies on the distinction between a mathematical "exercise" and a mathematical "problem"; or, between a "routine" problem and a "non-routine" problem. For the former, the solution method is generally known in advance. For example, many people who have completed courses in algebra would be able to solve the equation $2x - 3 = 5$. The process or algorithm for doing so is known in advance, and therefore, it is, in essence, just a matter of applying it. In contrast, a non-routine problem, or a mathematical problem in the formal, modern sense, is that for which the solution method is not known in advance (Schoenfeld, 1985, 1992, 2013). This conception of a mathematical problem aligns with the general problem definition given in psychological research literature. Dominowski & Bourne (1994) note that "Throughout the history of psychology there has been reasonable agreement that the essential features of a problem are that an organism has a goal but lacks a clear or well-learned route to the goal" (p. 23). Moreover, to solve a problem in this sense, one must overcome a blockage (McLeod, 1988) or lack of insight to attain the eventual solution that cannot be at the onset be overcome by previous knowledge or skills.

Research and materials related to the solving of such problems and generating solutions continued to increase steadily in the second half of the twentieth century (e.g., Bruner, 1962, 1966; Gagné, 1965; Krutetskii, 1969; Lakatos, 1976; Maier, 1970; Polya & Kilpatrick, 1974; Wason & Johnson-Laird, 1972; Wertheimer, 1945; Wickelgren, 1974); however, it was not until 1985 that Alan Schoenfeld published his seminal text *Mathematical Problem Solving.* In addition to providing an analytical synthesis of previous research, Schoenfeld synthesized the findings from several mixed-methods studies he conducted of mathematical problem solving behaviors and classroom practice. Schoenfeld conducted small-scale experimental studies of mathematical and cognitive task performance and observational studies of classroom practice. Based upon the results of these, Schoenfeld argued that there are four categories or dimensions of "knowledge and behavior that must be dealt with, if one wishes to 'explain' human problem-solving behavior" (Schoenfeld, 1985, p. 12). These are: resources (funds of available mathematical knowledge; understandings of the norms of mathematics and its discourse; relevant, discipline-specific competencies; etc.), heuristics (general problem solving techniques), control or metacognitive control (regulation of one's cognitive processes), and belief systems (one's own view of mathematics and themselves as mathematics learners).

2.3 Social and Cognitive Aspects of Problem Solving

The first two dimensions of Schoenfeld's framework (resources and heuristics) have been most naturally associated with perceived success in solving mathematical problems. Informally, one may expect that the more mathematical concepts one has mastered, and the more techniques they have developed to apply these concepts, the better. As Schoenfeld (2013) writes, mathematical knowledge and problem solving strategies form the basis of “what the individual has to work with” (p. 18). However, the dimensions of control (more recently referred to as metacognitive control) and belief systems have been the subject of a great deal of recent attention. For the latter, it has been extensively demonstrated that students’ beliefs about mathematics, about themselves as mathematics learners, and other affective variables (e.g., emotions) play a large role in successfully learning mathematics and solving problems (Gómez-Chacón, 2000, 2014, 2017; McLeod, 1988; Schoenfeld, 1985, 1988, 1989, 2013, 2016). In his review of literature on the social dimension of mathematics learning, Schoenfeld (2016) summarizes the viewpoint that mathematics is “an inherently social (as well as cognitive) activity, and an essentially constructive activity instead of an absorptive one” (p. 7). With respect to the development of learning and problem solving habits, Schoenfeld (1988) has written that “students learn what mathematics is all about by immersion in the routine, day-to-day practices of their mathematics classrooms” (p. 85). Moreover, the development of students’ beliefs about mathematics, problem solving, and their ability to be successful in these endeavors is to a large extent constructed socially.

Research has also indicated that the relationship between the social and the mathematical is reciprocal in nature. Although much research has been conducted on the influence of the nature and quality of social experiences on learning mathematics, several investigations have suggested that the nature and quality of mathematics learning influences important social competencies. In detailing findings from several educational studies, Ben-Avie et al. (2003a, 2003b) note consistent links between mathematical outcomes and social and developmental outcomes. They describe that mathematical problem solving is meaningfully related to social problem solving, goal-directed behavior, self-efficacy, and flexible thinking.

Research has also revealed that cognitive processes such as metacognition (one's own regulation of cognitive processes), self-regulation (active control over emotion, motivation, and cognition), and executive functioning (processes responsible for working memory, focus, action, and effort) are among the most important to the endeavor of solving mathematical problems (Schoenfeld, 2016). Research on metacognition "has evolved, according to Stillman and Mevarech (2010), from a 'hot topic' to a more 'mature field'" (Desoete & Craene, 2019, p. 566). As assessments of metacognition have improved in the nearly fifty years since the coining of the term, so, too, has the understanding of its role in mathematics problem solving. In their comprehensive review of metacognition research, Desoete & Craene (2019) argue that "metacognition seems to be one of the most important predictors of mathematical performance" (p. 565), and that decades of research support this stance. Desoete & Craene (2019) also note, as many researchers have, the propensity for metacognition to serve as an "umbrella term" (Schoenfeld, 2016, p. 22) for the cognitive processes related to mathematics learning and problem solving. Despite a lack of a formal, agreed-upon definition in the field, there is an emergent distinction between the recognition of one's own thinking, and the regulation of it. The latter is often referred to as "self-regulation" or the ability to exercise a level of "control" (to use Schoenfeld's words) over one's cognitive processes.

Research on self-regulation has also suffered from the same definitional problems as metacognition has, making the formation of a formal, agreed-upon definition equally difficult. Desoete & Craene (2019) note this difficulty in describing "the overlap between metacognition and self-regulation (executive functions)" (p. 568). Their parenthetical definition of self-regulation as consisting of the executive functions is a common one, and the research on the influence of these functions on mathematics learning and doing has become extensive in the last several decades. There is a general consensus that the executive functions comprise "*inhibition:* suppressing distracting information and unwanted responses, *shifting:* flexibility switching between different tasks, and *updating or working memory*: monitoring and manipulating information in the mind" (Cragg & Gilmore, 2014, p. 64). In their review of research on the role of executive functions on mathematics proficiency, Cragg & Gilmore (2014) found a strong, growing body of evidence suggesting a direct link between executive functioning skills and factual, procedural, and conceptual mathematical knowledge.

The review provided by Cragg & Gilmore (2014) focused on research that linked executive functioning skills and mathematical knowledge in children; however, the relationship between the two seems to be stable across age groups and into adulthood. Cragg et al. (2017) conducted an empirical investigation of the influences of executive functioning skills on mathematics achievement in individuals aged eight years to twenty-five. Cragg and her colleagues administered a battery of instruments measuring executive functioning, general mathematics achievement, and specific components of mathematics and learning. They determined that the executive functioning skills of working memory and inhibition were either directly or indirectly linked to mathematics knowledge (in this case, whole number arithmetic) and proficiency, and that this influence was stable across all age groups.

2.4 Inhibitory Control and Dual-Process Theories of Cognition

The role of inhibition, in particular, in mathematics learning and problem solving has been the subject of a great deal of recent research. In surveying the field of such research, Van

Dooren and Inglis (2015) note that “a great deal of theoretical work in mathematics education research seems to reference the notion of inhibitory control, albeit not always explicitly” (p. 713). Defined broadly as “the ability to ignore salient but unhelpful stimuli and responses” (Van Dooren & Inglis, 2015, p. 713), Van Dooren and Inglis note that decades of research on inhibitory control (IC) in mathematics suggests that it is a core cognitive construct influencing mathematics learning and problem solving.

One famous mathematics word problem that illustrates the necessity of IC to ignore unhelpful stimuli and intuitive, yet faulty reasoning is the famous “ball and bat” problem from the Cognitive Reflection Test (explained further in a later section). The problem states: “A bat and a ball cost $\$1.10$. The bat costs $\$1.00$ more than the ball. How much does the ball cost?” (Frederick, 2005, p. 26). This seemingly simple problem evokes the incorrect yet remarkably common answer of 10 cents or $\$.10$. Solvers will intuitively, and almost immediately, subtract $\$1.00$ from the $\$1.10$ to arrive at $\$.10$ as an answer, which reverses the “$\$1.00$ more…” part of the problem. This faulty, yet seemingly correct and immediate reaction is precisely that which needs to be inhibited in order to correctly answer the problem. If one engages their IC capacities and can inhibit this immediate reaction (essentially stopping and thinking before responding) they provide themselves the opportunity to engage in active, reflective thought about the task at hand. Research has shown that if this is the case, one can more readily arrive at the correct answer of $\$.05$ (since $\$.05 + \$1.05 = \$.05 + (\$.05 + \$1.00) = \$1.10$).

Building upon the early work of intuition in mathematics and science (perhaps most notably Fishbein, 1987), recent research has found direct links between inhibitory capacities and success in these areas. With respect to the role of IC as an executive function, Van Dooren & Inglis (2015) cite Miyake et al. (2000) in noting that some existing research suggests that inhibition may underly and "unify" the three executive functions, "as all executive functions involve some inhibitory processes to function properly" (p. 715). This may help contextualize research findings on working memory and executive functioning (e.g., the aforementioned Cragg et al., 2017) and contribute to a more comprehensive view of the role of the executive functions in mathematics learning and doing (for more on this, Van Dooren & Inglis point to Miyake et al., 2000 and Barkley, 1997).

The theoretical foundations for IC research in mathematics education can be attributed to the psychological research of higher cognition. Van Dooren and Inglis (2015) point to IC's central role in the prominent dual-process theory of cognition, which they note has accounted "for a great many results in the psychology of reasoning and decision making literatures" [1] (p. 717). Evans and Stanovich (2013) describe the theory as positing that cognition consists of two processes, Type 1 and Type 2. Type 1 processes are autonomous, immediate, intuitive processes triggered by stimuli that do not require controlled attention; and Type 2 processes are of higher order and are more reflective, effortful, and conscious (Evans & Stanovich, 2013; Leron & Hazzan, 2006, 2009; Stanovich & West, 2000). IC is central to this theory since one must inhibit Type 1 processes in order to engage in Type 2 processes. Evans and Stanovich (2013) write that "rapid autonomous processes (Type 1) are assumed to yield default responses unless intervened on by distinctive higher order reasoning processes (Type 2)" (p. 223). It is the immediacy of Type 1 processes and their independence of reflective thought can lead to errors in mathematical problem solving situations (Leron & Hazzan, 2006; Lubin et al., 2013; Verschaffel et al., 2020).

The defining characteristic that differentiates the two processes is working memory—simply put, "a process can be classified as Type 1 if it does not require working memory

[1] The authors point to Kahneman (2011) in this quotation for a "popular account" of such literature.

resources, and as Type 2 if it does" (Van Dooren and Inglis, 2015, p. 717). In the famous "ball and bat" problem described earlier, the Type 1 process being activated is the immediate, intuitive subtraction of $1.00 from $1.10 based upon the wording of the problem. This Type 1 response, if uninhibited, leaves the solver with the incorrect answer of $.10 without requiring any further reflection or use of working memory resources—the solver can simply move on. However, if this immediate reaction to respond with the intuitive response is inhibited, the solver is forced to activate working memory resources and engage in Type 2 processes as a result.

In the nearly fifty years since the origin of its first formulations, the dual-process theories of cognition have become a foundational component of not only psychological research, but also sociological and educational research. However, it should be noted, as Evans and Stanovich (2013) have, that "as the popularity of dual-process and dual-system theories has increased, so too have the voices of criticism" (p. 224). In their (2013) paper, Evans and Stanovich provide a detailed address of five major criticisms that have been leveraged against the theory in recent years, and address each in an analytical manner. Although Evans and Stanovich agree that the research on dual-process theories has created definitional problems in the field, they dispute the other claims made by critics of the theory (for these, Evans & Stanovich point to Keren & Schul, 2009; Kruglanski & Gigerenzer, 2011; Newstead, 2000; Osman, 2004). In essence, they argue that the major critiques of the theories arise from logical inconsistencies and semantic debates rather than from substantive evidence.

2.5 Measuring Inhibitory Control in Mathematics Education Research

Although the dual-process theories of cognition may be debated, what is clear from the extant literature is their utilization in explaining a wide range of psychological and educational phenomena. In particular, the application of these theories to mathematics education has aided in the understanding of how cognitive psychological variables influence mathematical outcomes (Gillard et al., 2009a; Gómez-Chacón, 2014; Van Dooren & Inglis, 2015). Glimore et al. (2015) note that studies investigating inhibitory control in mathematics utilize “a wide range of tasks to assess inhibition” and that these “tap into varying aspects of inhibition skill” (p. 772). They continue to note that a distinction exists between response inhibition and interference control tasks, and each can be administered in numerical and non-numerical formats. Less commonly used in the literature concerning the relationship between inhibitory control and mathematical knowledge are response inhibition tasks. In explaining response inhibition, Gilmore et al. (2015) note that:

> This concerns the suppression of a prepotent motor response and is often measured using Go/No-Go or Stop-signal tasks. In these tasks participants are required to frequently make one type of response unless they receive a signal to withhold the response. Inhibitory control performance is indexed by a failure to withhold the prepotent response. (p. 772)

Response inhibition tasks have been used in studies investigating the relationship between inhibition and mathematical outcomes in children with learning disabilities (e.g., De Weerdt et al., 2013). In the study conducted by De Weerdt et al. (2013), the participants were administered response inhibition tasks on a computer, where participants were instructed that a one group of images that may appear are “go” prompts and another group were “no go” prompts. When the “go” prompts were presented on the screen, the participants were to react by pressing a key on the keyboard; when the “no go” prompts were presented, they were not to react.

In contrast, Gilmore et al. (2015) explain that interference control tasks require “the suppression of distracting information, either internal or external, which leads to an alternative non-desired response" (p. 772). Perhaps the most famous such task is the Stroop task (Stroop, 1935) which has been replicated, modified, and extended many times in both numerical and non-numerical formats. For example, in their investigation of inhibitory control and component arithmetic knowledge, Gilmore et al. (2015) utilized both non-numerical and numerical versions of interference control tasks based upon the Stroop task. In the non-numerical inhibition task, participants were required to view two animals and decide which was physically larger in real life. However, the sizes of the images were manipulated in each trial; for example, in one trial, an image of a butterfly was presented as much larger than an image of an elephant. The participant would need to inhibit the reaction of choosing the animal that is depicted as being larger; in the aforementioned example, the participant would need to choose the elephant even though it is shown as being smaller than the butterfly. Gilmore et al. (2015) also utilized a numerical inhibition task based on the Stroop task which required participants to determine which of two presented images contained more dots. The sizes of the dots were manipulated on different trials to encompass more area on the screen. The participant would need to inhibit the reaction of just looking at area encompassed by the dots to make a decision, and instead actually count the dots themselves.

The "ball and bat" problem of the Cognitive Reflection Test (discussed in a preceding section) is another task that measures inhibitory control. Recall that the problem states, "A bat and a ball cost $1.10. The bat costs $1.00 more than the ball. How much does the ball cost?" (Frederick, 2005, p. 26). In comparing the tasks just described with the ball and bat problem, Van Dooren & Inglis (2015) write that there "substantial difference between the tasks" (p. 714) and explain that

> For instance, the bat and ball problem seems to be substantially more demanding than the numerical Stroop task. Whereas the latter requires the processing of a physical property to be inhibited, the Cognitive Reflection Test is entirely cognitive: a salient response must be inhibited and some analytical work done before a correct answer can be given. (p. 714)

It is important to note that mathematical word problems used to measure inhibitory control, like the "ball and bat" problem, incorporate language in a way that formations of the Stroop task and Go/No-Go or Stop-signal tasks do not. The phrase "costs $1.00 more than the ball" immediately evokes the mathematical operation of addition, which in turn evokes the reverse of that operation to solve (incorrectly) the problem. Other studies have utilized tasks that rely on the language of the problem as a core component. The negative priming (NP) paradigm (see Tipper, 1985, 2001, and Neill et al., 1995 for more on the history and development of NP research) has recently been utilized by Lubin and colleagues (2013, 2016) to investigate the relationship between inhibitory control and word problem solving performance.

Lubin et al. (2016) note that "The NP paradigm rests on the basic principle that if you inhibit a given strategy on one trial, then the activation of this strategy on the next trial should be more difficult" (p. 41). In essence, the response time and less accurate responses given on subsequent (called "priming") tasks provides an indication of this difficulty, and can be directly measured. One subset of these problems are inconsistency language (IL) problems, where the language of the problem is in conflict with a previously-learned yet faulty heuristic (that needs to be inhibited) for solving the problem. For example, one traditional heuristic is the association of the operation of addition with language such as "more" or "more than" and subtraction with language such as "less" or "less than". Although this heuristic is developed by success on, and used to solve consistent language (CL) problems, this heuristic needs to be inhibited to be successful with IL problems. The NP paradigm has also been used as components of more comprehensive studies investigating inhibitory control and dual-process theories of cognition. For example, Houdé & Borst (2014) further investigated cognitive inhibition using both the NP paradigm alongside brain images of the prefrontal cortex using an fMRI (functional magnetic resonance imaging).

Another trend in the measurement of inhibitory control in mathematics education literature is one that focuses on types of mathematical intuitions. In certain situations, well-learned heuristics or intuitions about the nature of mathematics or its properties need to be inhibited in order to successfully engage with a mathematical task or problem. In the introduction to their (2015) paper, Attridge and Inglis (2015) detail five of the major intuitive conflicts that arise in mathematical situations: the natural number bias, proportional reasoning, geometric reasoning (areas and perimeters), probability, and the famous "students and professors problem" (Clement et al., 1981). The first three of these seem to be the most frequently used tasks in recent literature investigating inhibitory control and mathematics knowledge.

In the case of the natural number bias, it has been shown that children and adults are affected at varying levels of mathematical expertise by their knowledge of natural numbers in dealing with rational numbers (Obersteiner et al., 2013, 2016; Vamvakoussi et al., 2012; Van Hoof et al., 2013). An important feature of natural number bias tasks (as well as in in proportional, geometric, and probabilistic reasoning tasks) is the presence of "congruent" and "incongruent" items (the pairing of these items is similar to that of pairing consistency language word problems and inconsistency language word problems as discussed earlier). In congruent items, the correct answer or solution to the task is aligned with the learned heuristic of natural numbers. For example, one task given in Vamvakoussi et al. (2012) requires a true or false response and is stated “$10y + 1$ is always greater than $10y$” (p. 326). This is a true statement and is aligned with the knowledge that addition with natural numbers always results in a quantity of greater magnitude. In contrast, a statement such as “$5 \div x$ can be bigger than 5” (p. 326) is true, and is an incongruent task since it requires the inhibition of the learned natural number heuristic that division always results in a quantity of lesser magnitude.

Similarly, in proportional reasoning tasks, congruent (or “proportional”) items are those that are aligned with proportional reasoning heuristics, and incongruent (or “nonproportional”) items are those where this heuristic needs to be inhibited. Take for example the following tasks given by Gillard et al. (2009b),

> Proportional:
>
> Erik and Tom buy boxes of pencils in the shop. All boxes are equally expensive, but Erik buys fewer boxes. Erik buys 4 boxes of pencils, while Tom buys 8 boxes. Knowing that Erik has to pay 24 Euros, how much does Tom have to pay?
>
> Nonproportional:
>
> Ellen and Kim are running around a track. They run equally fast but Ellen started later. When Ellen has run 5 laps, Kim has run 15 laps. When Ellen has run 30 laps, how many has Kim run?
>
> (p. 94)

In the first problem, it can be determined that each box of pencils costs 6 Euros ($24 \div 4$). Then, using proportional reasoning, it must be the case that Tom has to pay 48 Euros (since $6 \times 8 = 48$). In the second problem, as Attridge and Inglis (2015) note, since Kim has originally ran three times as many laps as Ellen, the incorrect yet common solution (which leverages proportional reasoning) would be 90 (30 laps times 3). However, the correct answer is 40 based upon the fact that both runners complete laps at the same rate, and one just began earlier. In geometric reasoning tasks, congruent items are aligned with the learned geometric heuristic that the more area a figure has, the larger its perimeter; in incongruent tasks, this heuristic must be inhibited since it breaks down in certain situations (see Babai et al., 2010, 2015 and Stavy & Babai, 2010).

2.6 Results from Inhibitory Control Studies in Mathematics Education

A substantial amount of research on the relationship between inhibitory control and academic outcomes in mathematics has been conducted on young children since inhibitory control is one of the core executive functions, which as Diamond (2013) notes "are skills essential for mental and physical health; success in school and in life; and cognitive, social, and psychological development" (p. 136). Thus, many studies that investigate children's development incorporate measures of the executive functions in the prediction of academic outcomes, particularly in the domains of reading and mathematics in preschool, kindergarten, and elementary school (Diamond, 2013 points to Blair & Razza, 2007; Borella et al., 2010; Duncan et al., 2007; Gathercole et al., 2004; Morrison et al., 2010; and Spiegel, 2021). Links between inhibitory control and early academic outcomes, particularly in mathematics, have also been found (Allan et al., 2014; Bull & Lee, 2014; Clark et al., 2010; Jacob & Parkinson, 2015; Son et al., 2019).

As Coulanges et al. (2021) notes, in comparison to the research conducted on executive functioning in mathematics on students from the early age groups, "far less is known about the contributions of these skills to the mathematical content taught in secondary school and even less at the college level" (p. 11). However, several investigations have indicated that inhibitory control is an important cognitive ability influencing success on mathematical tasks for students throughout adolescence through the middle and secondary grades (Babai et al., 2010, 2015; Gómez-Chacón et al., 2014; Van Hoof et al., 2013; Van Dooren et al., 2005). Recent research has also investigated whether inhibition skill was related to procedural, conceptual, and factual arithmetic knowledge, and whether this relationship was consistent from childhood to adulthood (Gilmore et al., 2015). In alignment with previous research, statistically significant relationships were found by Gilmore et al. (2015) between inhibition skill and arithmetic knowledge across all age groups. The researchers also found that inhibition skill was linked to procedural knowledge in children and conceptual knowledge in adults, the relationship between inhibitory control and mathematics knowledge may be specific in nature, and that this relationship may evolve with age.

Lubin et al. (2013) utilized the NP paradigm to investigate the relationship between children's and adult's inhibitory control skills and their performance on word problem solving tasks. The researchers administered inconsistency language (IL), consistency language (CL), and neutral verification (asking participants to verify problem solutions) word problems to sixth-graders, ninth-graders, and adult college students. The researchers found that although children were less efficient than adults at solving IL problems (a finding consistent with previous research), the necessity of inhibitory control to solve these problems was consistent across all age groups. This suggests that inhibitory control as an explanatory variable for success on mathematical tasks is stable over time but may be subject to differences in age since individuals mature cognitively as they age.

Recent research has also explored the influence of mathematical expertise on success on mathematical tasks that require inhibitory control. Lubin et al. (2016) advanced their (2013) work to further investigate inhibitory control capacities in adults with respect to their performance on arithmetic word problems. The researchers also sought to investigate the potential influence on mathematical expertise on adults' inhibitory and word problem solving skills. Lubin and her colleagues conducted two studies which utilized the NP paradigm and administered control pairs of IL and CL problems to adult college students who were mathematical experts (mathematics undergraduates) and non-experts (undergraduate humanities students). The second of the two studies also accounted for differences in general intelligence and inhibitory control ability. The analyses from both experiments suggest that although experts in mathematics still need to inhibit faulty heuristics when solving these problems, they are more efficient than non-experts at doing so. Based upon the analyses from the second experiment, this efficiency can be attributed to a "specific ability to inhibit a misleading heuristic" (p. 46) rather than differences in general intelligence and IC ability.

Taken together, the results from Lubin et al. (2013, 2016) seem to indicate that although mathematical experts and adults tend to perform better on tasks that require the ability to inhibit faulty responses and heuristics than non-experts and children, respectively, this ability is required by individuals of all ages and varying levels of expertise in order to accurately solve mathematics word problems. Research has also shown that adults, including college students, mathematics undergraduates, preservice teachers, and mathematicians are affected by mathematical intuitions such as the natural number bias and experience difficulty activating inhibitory control mechanisms in order to successfully engage with mathematical tasks (Coulanges et al., 2021; Gillard et al., 2009; Houdé & Borst, 2014, 2015; Obersteiner et al., 2016; Rossi et al., 2019; Vamakoussi et al., 2013).

2.7 The Cognitive Reflection Test

Research on relationship between inhibitory control capacities in adults and mathematical outcomes have increased substantially in the last several decades. The techniques and instruments described in an earlier section have also experienced a great deal of development; however, few have rivaled the popularity of the Cognitive Reflection Test (CRT) developed by Frederick (2005). Young and Shtulman (2020) even note that the CRT “is the dominant measure of adult individual differences in analytic versus intuitive thinking” (p. 1396). The test consists of three mathematical word problems that are “hard to solve, not because of their linguistic or mathematical demands but rather because the correct answer requires the inhibition of a very easy, seemingly correct answer” (Verschaffel et al., 2020, p. 9). Therefore, “cognitive reflection” in this sense refers to the ability to inhibit an intuitive response to a stimulus (i.e., IC), and engage in “reflective” or active thought. Thus, in essence, the CRT is an instrument that measures IC ability, since IC is required to correctly solve each problem.

Interpreted through the dual-process theories of cognition, the CRT measures one’s ability to inhibit immediate, intuitive, or default responses to a stimulus (Type 1 processes) with the Type 2 processes that are more reflective, conscious, and which use working memory resources. Arguably the most famous task on the CRT is the “ball and bat problem” discussed several times throughout this review, which has been characterized as a “classic dual-process default-interventionist task” (Van Dooren & Inglis, 2015, p. 717). In the sense described earlier in this section, the second and third problems of the CRT can be characterized as a type of modified proportional reasoning inhibitory control tasks. The problems of the CRT are given below:

1. A bat and a ball cost $1.10. The bat costs $1.00 more than the ball. How much does the ball cost? ______ cents
2. If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? ______minutes
3. In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake? _______days

(Frederick, 2005, p. 27)

In his original study, Frederick (2005) compiled thirty-five separate empirical investigations involving 3,428 conveniently-sampled and incentivized participants, most of whom were college students. The participants were administered the CRT and a battery of other instruments measuring general intelligence, risk preferences, time preferences, and decision-making. Correlational analyses were conducted between measures on each of these instruments across different subgroups (gender, college students from different universities, etc.). Performance on the CRT was found across subgroups to be linked to differences in these measures. With respect to intelligence, the results suggested that CRT performance was not solely explained by measures of general intelligence (e.g., the Wonderlic Personnel test, the SAT/ACT, the Need for Cognition Scale), but is positively influenced by reading comprehension and mathematics proficiency. These findings lend some support to claims that the CRT measures a cognitive capacity distinct from general intelligence.

Although Frederick's original study investigated the relationships between CRT performance and general cognitive and decision-making outcomes, it has since been incorporated into a wide-range of studies investigating a broad set of psychological and behavioral outcomes.

Such research on the CRT since 2005 has included investigations of alternate forms of the instrument (Otero, 2019; Primi et al., 2015; Sirota et al., 2021; Thomson & Oppenheimer, 2016; Young & Shtulman, 2020); moral judgement (Baron et al., 2015); general intelligence and numerical ability (Liberali et al., 2012; Otero et al., 2022; Welsh et al., 2013); mathematics anxiety (Jiang et al., 2021; Morsanyi et al., 2014); thinking and heuristic biases (Stanovich & West, 2000; Toplak et al., 2011); biases and fallacies in probability judgement (Liberali et al., 2012); information processing (Toplak et al., 2014); and science and mathematical learning outcomes in college students and children (Shutlman & McCallum, 2016; Young & Shtulman, 2018).

Although decades of research has produced promising results in studies of the CRT and its alternate forms in explaining a wide range of psychological phenomena, there also exists evidence that suggests that the CRT has become less reliable due its popularity and correlations to measures of general intelligence. The immense popularity of the CRT has caused some researchers to consider whether the CRT has become a “victim of its own success” (Haigh, 2016, p. 145). Haigh (2016) and others (see also Stieger & Reips, 2016) have found that prior exposure to the test is highly correlated with success on future administrations, which calls into question the validity of many CRT-related findings. Several empirical studies (e.g., Welsh et al., 2013) and several large-scale meta-analyses (e.g., Otero et al., 2022) have also called into question the CRT’s validity in measuring a distinct cognitive construct rather than just another instantiation of general intelligence.

Despite the findings of many critics of the CRT, a growing body of research continues to suggest that the CRT is a legitimate measure of inhibitory control/cognitive reflection, is stable over time, and that prior exposures to the instrument do not have a significant effect on future performance. For example, Stagnaro et al. (2018) conducted eleven separate studies of 3,302 participants recruited from Amazon Mechanical Turk who have previously completed the CRT two or more times. In contrast to the research of Haigh and others, Stagnaro and colleagues found that there was a high correlation observed between these participants' first and last CRT performance, suggesting that prior exposure to the test does not influence future performance (see also Bialek & Pennycook, 2017), and that such performance is stable over time. In summarizing their study investigating numeracy and cognitive reflection in college students, Liberali et al. (2012) note that "the answer to the question of whether CRT is just another numeracy test would seem to be a qualified no" (p. 379). In pursuit of isolating exactly what the CRT measures, Pennycook et al. (2015) summarize that their results "provide evidence that the CRT is a valid measure of reflective but not of intuitive thinking" (p. 341). Additionally, they point to several studies (Campitelli & Gerrans, 2014; Cokely & Kelley, 2009; Toplak et al., 2011, 2014) in noting that "multiple investigations have established that CRT performance is not fully explained by numeracy or cognitive ability" (p. 346).

Despite numerous studies investigating the link between the CRT and general numerical ability and cognitive intelligence, the majority of these attempt to address numerical ability through the use of standardized assessment scores (e.g., the SAT/ACT) or numeracy scales (e.g., the numeracy scales of Fagerlin et al., 2007; Lipkus et al., 2001; Schwartz et al., 1997). Previous research has indicated that the CRT is more difficult than previously used numeracy scales (Weller et al., 2012) and measures a distinct cognitive construct separate from numeracy (Liberali et al., 2011). However, specific investigations of mathematical expertise accounted for by different measures, specifically as measured by the extent of mathematical learning in the classroom, are far less common. It has been shown that university students perform better on the

CRT than non-university students (Brañas-Garza et al., 2019) and that students from different universities perform differently (Frederick, 2005); however, more research that investigates the performance of college students from different majors and levels of mathematical expertise is needed.

Mathematics experts (graduate students in mathematics and professional mathematicians) have been shown to perform better on inhibition tasks and are less susceptible to faulty intuitions and heuristics than non-experts (Houdé & Borst, 2014; Kahneman, 2011; Vamvakoussi et al., 2013; Orbersteiner et al., 2013, 2015; Lubin et al., 2016); however, an investigation of university students learning mathematics at different rigor and complexity levels in relation to measures of inhibition, especially the CRT, is needed to add to this literature. Moreover, the differences in mathematical knowledge accounted for by the measured success in university mathematical coursework has yet to be extensively investigated in relation to cognitive reflection as measured by the CRT.

The only study known to the author to investigate measures of success in mathematical coursework in relation to the CRT was conducted by Gómez-Chacón et al. (2014) in their investigation of secondary school students. The researchers found that CRT performance was positively correlated with mathematics performance in the course in which they were enrolled. This study was also one of the only known studies to study the relationship between responses to a questionnaire of mathematics beliefs and CRT performance. The researchers also found positive beliefs in mathematics were correlated with CRT performance. Such an investigation that investigates the relationship between mathematics coursework, CRT performance, and belief systems in college students would meaningfully contribute to extant literature.

2.8 Learning and Development in College Students

Also yet to be conducted is an investigation of whether CRT measures are linked to other psychoeducational variables that are known to influence college students' academic success. As previously mentioned, a great deal of research on inhibition has been conducted on younger students, particularly those in the primary and middle grades, with far less being conducted on students in secondary and postsecondary schools. Research of this type on students in the early grades has also explored the relationships between inhibition and a wide range of other psychological variables. Research on the CRT has attempted to do this, following the aims of Frederick's (2005) initial study, where the CRT was administered alongside other psychological and cognitive assessments. However, investigations into the potential relationships between CRT performance and psychological and educational variables in college students are far less common. Although a robust explanatory framework for the learning and development of college students exists, it has not yet been extensively explored in relation to the CRT.

With respect to learning and development, over half a century of educational research has established "development and learning are inextricably linked, but traditionally development is not intentionally addressed" (Comer, 2022). The work of James P. Comer and his colleagues has demonstrated that schoolchildren's' psychological, cognitive, physical, linguistic, social, and ethical development is inextricably linked with their academic learning outcomes (Comer 1992 1993, 2005; Comer et al., 1999; Brown and Corbin, 2004). In addition to serving as a framework for school reform, this research has demonstrated that in order to help children reach their potential, all stakeholders in the educational environment must work "with the whole child in mind" (Darling-Hammond et al., 2018).

Research conducted by Michael Ben-Avie and his colleagues has demonstrated that the same developmental experiences that are inextricably linked with learning outcomes in school children are so in college students, which has contributed to a robust explanatory framework for college student success (Ben-Avie, 2008, 2013, 2018; Ben-Avie & Darrow, 2018, 2019; Ben-Avie et al., 2012; Ben-Avie & Polka, 2006; Darrow, 2016, 2020; Pang et al., 2016). Longitudinal cohort analyses of thousands of college students have consistently shown that facets of college students' psychoeducational profile, such as their academic habits of mind, future orientation, academic self-concept, self-limiting beliefs, and sense of belonging, meaningfully predict their academic success, persistence, and retention in college (Ben-Avie et al., 2012; Ben-Avie & Darrow, 2018, 2019).

Academic habits of mind include the skills, attitudes, and the regulation of these in the academic environment as well as "self-regulation, competency to work autonomously, time management, study skills, and the process of inquiry that is common to all academic disciplines" (Ben-Avie & Darrow, 2019, p. 23). Ben-Avie & Darrow (2019) define future orientation as "the ability to conceive of one's own development and take actions in the here-and-now to achieve one's hoped-for future (Ben-Avie et al., 2003)" (p. 23). Future orientation is generally a positive attribute; however, maladaptive aspects to future orientation exist as well. For example, students who are solely focused on the future, or those who are worried about the future and potential negative events to come, can lose sight of their present time perspective and current focus. Academic self-concept is one's view of themselves within the academic domain. More generally, Chiu & Klassen (2010) describe self-concept in their review of literature:

> Self-concept is defined as self-perceptions about one's abilities and competences (Byrne & Shavelson, 1986) that influence the likelihood of success in a wide range of endeavours. When people have positive self-concept, they show more motivated

> behaviors and greater perseverance with challenging tasks (Stipek, 1998). However, self-concept also involves feelings one has about oneself (i.e., self-esteem; Harter, 1985) as well as self-efficacy and perceptions of others' responses to one's self (Dermitzaki & Efklides, 2000); that is, self-concept consists of self-beliefs that are formed through interaction with the environment. Self-concept is a multidimensional, hierarchical construct that is influenced by social comparison, causal attribution, appraisals from significant others, and mastery experiences (Bong & Skaalvik, 2003). (p. 3)

Students' academic self-concept is related to their academic belief systems, which can be discipline specific. Schoenfeld (1985) notes that students belief systems in mathematics consist of "one's 'mathematical world view,' the set of (not necessarily conscious) determinants of an individual's behavior about self, about the environment, about the topic, about mathematics" (p. 15). Self-limiting beliefs in mathematics (e.g., "I'm just not good at math") are also related to one's academic self-concept in the discipline. The interactions between self-limiting beliefs, academic habits, academic self-concept, and future orientations should also be noted. For example, self-limiting beliefs have the potential to subvert future orientation (Ben-Avie & Darrow, 2019; Dilts, 1999) and perpetuate negative feelings such as hopelessness, helplessness, and worthlessness (Dilts, 1999, p. 116), which can negatively influence one's self-concept and the quality of one's academic habits.

To investigate the influence of such factors on college students' learning and development, Ben-Avie & Darrow (2019) conducted multiple longitudinal cohort studies of several thousand students spanning more than ten years. The researchers followed several cohorts of thousands of students' academic paths throughout college in their investigation of two types of characteristics of a college students' psychoeducational profile: immutable characteristics and malleable characteristics. They note that immutable characteristics “include high school records, precollegiate developmental experiences, ethnicity, and socioeconomic status during childhood” and are “unalterable or are not amenable to change after enrollment in college” (p. 22). Conversely, malleable characteristics, such as academic commitment and habits of mind, future orientation, academic self-concept, self-limiting beliefs, and sense of belonging can be influenced while students are in college. Much of the research on predicting collegiate academic success over the past forty years has largely focused on the predictive power of immutable characteristics; however, Ben-Avie & Darrow (2019) found that in comparison, malleable characteristics were stronger, more consistent predictors of overall academic success, retention, persistence, and graduation than immutable ones.

Although such factors of a students’ psychoeducational profile have been traditionally difficult to measure, “these factors are directly related to teaching and learning in and out of the classroom” (Ben-Avie & Darrow, 2019, p. 47). In order to measure these factors, Ben-Avie and his colleagues (Ben-Avie et al., 2012; Ben-Avie & Darrow, 2019) developed and refined psychometric instruments to reliably measure academic habits of mind, future orientation, academic self-concept, and self-limiting beliefs (see Ben-Avie et al., 2012 for a description of the psychometric properties of these instruments). Examples of Likert scale items (five point: strongly disagree, disagree, neutral, agree, strongly agree) from these instruments that measure academic habits of mind are given below (these items were adapted from the work of Ben-Avie & Darrow, 2019 and are included in the Short Mathematics Psychoeducational Inventory appearing in Appendix B of this document):

Academic Habits of Mind:

- I study regularly to be successful in college.
- I have different "game plans" for tackling homework assignments, and I know which strategy would be the most effective for each type of assignment.
- I settle for just passing my courses.

Future Orientation:

- I have a fairly clear idea of what I need to study now in order to have the career I want.
- Thinking about the future I want makes me do more now to get that future.
- Thanks to this university, I have a clear idea of what I need to study in order to have the career that I want.

Self-Limiting Beliefs

- I fear that if I ask for help, my professor will think less of me.
- I am hesitant to raise my hand in class even when I know the answer.
- I'm just not good at math.

Academic Self-Concept

- I am doing better than I thought I would in college.
- Sometimes, I am disappointed in my test results because I studied a great deal.
- My confidence in academic skills has increased this semester.

(Appendix B, this document)

These items have consistently been shown to measure the developmental components of a students' academic profile that cannot be gleaned from direct academic measures (e.g., grades, grade point average or GPA, standardized test scores, crystallized knowledge, etc.). Although the research has shown that malleable characteristics hold more predictive power than immutable ones, this does not mean that students' past experiences and demographic makeup are not important. Recall that research on the inextricable link between learning and development has demonstrated that educators must consider the “whole child”, or in this case the whole college student, since students’ past experiences and demographic characteristics are “pivotal in understanding students’ success in college” (Ben-Avie & Darrow, 2019, p. 28) and can lead to “understanding student backgrounds, experiences, and expectations so that institutions can minimize unmet expectations and increase student engagement, learning, satisfaction, and persistence” (Cole et al., 2009, p. 67).

In a longitudinal cohort investigation of mathematics placement methods, Darrow (2016) incorporated information on students’ past experiences alongside measures of malleable characteristics to predict success in students’ first mathematics course in college, and their overall academic success, persistence, and graduation. He found that measures of students’ precollege developmental experiences, developmental experiences on campus, high school grade point average (GPA), academic self-concept in mathematics, and the level and rigor of their high school mathematics courses were better predictors of success in their first mathematics courses than the traditionally used placement methods of standardized tests such as the SAT. Darrow also found that a consistent, positive relationship between desirable measures of psychoeducational variables (academic habits of mind, future orientation, etc.) and success in the mathematics classroom, particularly in students’ first mathematics class. Moreover, performance in students’ first mathematics courses in turn emerged as one of the most consistent, strong predictors of overall academic success, persistence, and graduation.

Building upon the work of Darrow (2016) and Ben-Avie and Darrow (2018, 2019), Darrow (2020) developed and administered a psychoeducational survey consisting of domain-general and mathematics-specific items measuring academic habits of mind, future orientation, academic self-concept, self-limiting beliefs, problem solving beliefs, and past academic experiences in mathematics to investigate mathematical belief systems and problem solving behaviors in college students. Since examples of domain-general items were given above, examples of mathematics-specific forms of these items are given below (taken from the Short Mathematics Psychoeducational Inventory appearing in Appendix B of this document):

Academic Habits of Mind in Mathematics

- I am always well-prepared for math class.
- I push aside math assignments and do them last.
- I wait until right before a math test to start studying.
- I have a ‘game plan’ that is effective for tackling math homework.
- I tell my professor when I don’t understand something from math class.

Past Academic Experiences in Mathematics

- Math and/or anything with numbers has been an obstacle to my academic success.
- My grades in college math have influenced what degree I can pursue.
- I have experienced difficulties in math since high school or before high school.

Academic Self-Concept (In Mathematics)

- I am usually confident that I will do well on math tests.

Problem Solving Beliefs

- The wording of math problems confuses me.
- The hardest part about solving word problems is understanding what is being asked.

(Appendix B, this document)

In addition to taking the survey, a subset of participants also recorded their thought processes and behaviors while solving mathematics problems. This subset also provided reflections on their past experiences in mathematics and how these have influenced their learning in college. The responses from the survey, contextualized through students’ authentic writing about their past experiences, belief systems in mathematics, and their academic self-concept in the discipline, revealed the importance of combining students' developmental experiences (including malleable characteristics of their academic profile) and past experiences in mathematics helped to explain their learning in college mathematics courses.

Mathematics-specific measures of self-concept and academic habits of mind (particularly whether students had an effective 'game plan' for engaging with mathematical work) emerged as particularly important in relating both students' experiences in mathematics as well as their problem solving behaviors. Moreover, the results seem to compliment findings from previous research (e.g. Ben-Avie & Darrow, 2019) which describe the importance of students' psychoeducational profile in explaining academic outcomes in mathematics and in college in general.

It should be noted that a number of studies have previously investigated scales and other measures of mathematics to measure students' (including college students) belief systems in mathematics (e.g., Gómez-Chacón et al., 2014; Hudson et al., 2012; Stage & Kloosterman, 1991; Kloosterman & Stage, 1992). Additionally, only one known study conducted by Gómez-Chacón et al. (2014) investigated these measures in relationship to CRT performance in high school students. However, those administered in the study by Darrow (2020) are new and distinctly different from these and those utilized in other studies in that they were constructed from existing psychoeducational items developed by Ben-Avie et al. (2012) and Ben-Avie and Darrow (2019). These mathematics-specific psychoeducational items were developed to measure indicators of academic habits of mind, self-limiting beliefs, academic self-concept, etc. (which have already shown to be related to collegiate academic outcomes) in college students as well as attributes that are amenable to change within the domain of mathematics. In other words, the theoretical background informing the construction of these items is based on this subset of research.

2.9 Toward the Current Study

In this chapter, a great deal of research investigating the cognitive, social, and psychological variables that contribute to success in mathematics learning and doing was reviewed. To provide further context for the aims of the current study, the following list summarizes the major themes discussed in this chapter:

- Although a great deal of problem solving research has been conducted, recent work has begun to provide insight into the sociological, psychological, and cognitive variables affecting mathematics problem solving.
- Using Schoenfeld's (1985) framework for problem solving as a basis, there are four main dimensions of problem solving: resources (funds of available mathematical knowledge; understandings of the norms of mathematics and its discourse; relevant, discipline-specific competencies; etc.), heuristics (general problem solving techniques), metacognitive control (regulation of one's cognitive processes), and belief systems (one's own view of mathematics and themselves as mathematics learners)
- Among the cognitive processes within the metacognitive component of problem solving are self-regulation and the executive functions (shifting, working memory, and inhibition or inhibitory control)
- Inhibitory control—or the ability to suppress immediate, intuitive reactions to a stimulus and engage in deeper, more reflective thought—has been shown to be related to success on a wide range of mathematical problem solving tasks

- Previous mathematics education research involving inhibitory control has utilized a wide range of instruments. Among these is the Cognitive Reflection Test (CRT), which has been widely administered to adult college students.
- Investigations of the influence of mathematical expertise on CRT performance have focused on measures of intelligence and numeracy scales. Investigations involving other instruments measuring inhibitory control have utilized different measures of mathematical expertise, such as the extent of mathematical coursework taken.
- Research on college students has shown that psychoeducational facets of college students' academic profiles meaningfully explain and predict learning and developmental experiences in college as well as collegiate academic outcomes. Recent research has attempted to extend this to the measuring of learning and developmental experiences specific to mathematics, however an extensive investigation of this has yet to be conducted.

The research discussed in this chapter was particularly chosen to provide a brief, yet informative review of existing literature related to mathematical problem solving; the cognitive, psychological, and sociological variables influencing mathematical problem solving; inhibitory control, its measurement, and its presence in mathematics education research; and the learning and development of college students. Moreover, in addition to the section focusing specifically on college students' learning and development, a particular attempt was made to highlight research conducted in college student populations relating to the other aforementioned areas. Therefore, although a great deal of research on these topics exist that was not discussed in this chapter, the review presented here provides a limited, yet functional framework for understanding the variables related to college students' learning, development, and mathematical problem solving.

An illustration of this framework is given below in Figure 2.9.1. This figure provides visual description of the relationships among the variables discussed in this review.

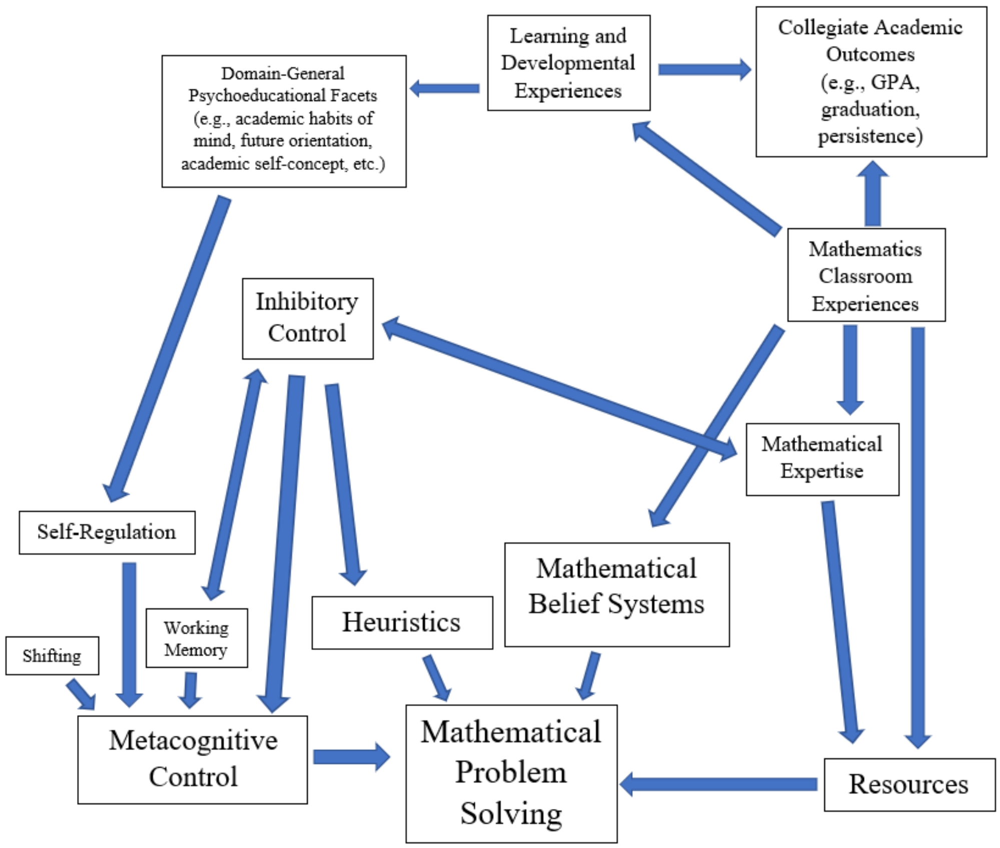

Figure 2.9.1: A Functional Framework of Theoretical Constructs Discussed in the Review of Literature

First, as seen in the diagram, the direct relationships between the four dimensions of problem solving from Schoenfeld's (1985) framework and mathematical problem solving are given (resources, heuristics, metacognitive control, and belief systems). Then, the relationships between several metacognitive components of problem solving are illustrated, including the executive functions of inhibitory control, shifting, and working memory as well as self-regulation. Also illustrated is the influence of mathematics classroom experiences on the development of mathematical expertise and belief systems in mathematics. Note also that mathematics classroom experiences and the development of mathematical expertise influence resources (funds of available mathematical knowledge; understandings of the norms of mathematics and its discourse; relevant, discipline-specific competencies; etc.). Additionally, mathematics classroom experiences influence both collegiate academic outcomes as well as the learning and developmental experiences in college students, which are known to be related to psychoeducational facets of students' academic profiles. The illustration also depicts inhibitory control and its reciprocal relationship to working memory, heuristics (inhibitory control is needed to inhibit the use of intuitive yet faulty heuristics), mathematical expertise, and the direct relationship between inhibitory control and metacognitive control.

In summary, the limited, yet functional framework detailed above provides a foundation from which to view the methods and results of the current study. This framework will be recalled in the next chapter and in the final chapter to further contextualize the contributions of the current study.

Chapter 3: Methodology

3.1 Research Questions

The purpose of this study is to explore the links between cognitive, social, and psychological variables that are known to influence collegiate academic outcomes through the administration of a modified version of the Cognitive Reflection Test and a psychoeducational survey to college students of differing levels of mathematical expertise. Specifically, this mixed-methods empirical study sought to answer the following research questions:

1. How do students from different collegiate mathematics courses perform on the Cognitive Reflection Test (modified by the investigator to include two additional problems)?
2. How do students from different collegiate mathematics courses respond to a psychoeducational survey measuring domain-general academic habits of mind, future orientation, and academic self-concept, and mathematics-specific academic habits of mind, beliefs, and past academic experiences?
3. What is the relationship between the performance of students from different collegiate mathematics courses on the modified version of the Cognitive Reflection Test and their responses to a psychoeducational survey measuring domain-general academic habits of mind, future orientation, and academic self-concept, and mathematics-specific academic habits of mind, beliefs, and past academic experiences?

3.1.1: Brief Methodological Overview

Before a detailed discussion of the methods of the current study, a brief overview of these and how they contribute to the answering of the aforementioned research questions is given below.

- To answer Research Question 1, a modified version of the Cognitive Reflection Test was administered to 130 college students, each differing with respect to their enrollment in different mathematics classes at the university under study. These classes, and thereby the students, were grouped into four subgroups according to the rigor and complexity of the content covered in these classes and the nature of prerequisite requirements needed to enroll in each course. This yielded a hierarchy of mathematics courses which served as a proxy measure of mathematical expertise. A descriptive and inferential analysis of the results of participants' performance on the modified version of the Cognitive Reflection Test was undertaken. The findings were further contextualized by additional subgroup information that was collected (e.g., gender, academic year, academic major, high school mathematics courses taken).
- To answer Research Question 2, a survey instrument, developed from previous research and consisting of items measuring domain-general and mathematics specific psychoeducational facets of students' academic profiles, was administered to all 130 students from different mathematics course level subgroups. An analysis of the instrument (which included factor analysis) was undertaken, and the results informed the development of reliable scales. Participants' responses to the items of these scales were descriptively and inferentially analyzed to determine mathematics course level and other subgroup differences. The findings were further contextualized by analyses of additional subgroups.
- To answer Research Question 3, participants' performance on the modified version of the Cognitive Reflection Test and responses to the survey instrument were analyzed descriptively and inferentially in relation to one another. The relationships between the

measures of both instruments were explored across the entire sample, each of the mathematics course level subgroups, and each of the additional subgroups under measure.

3.1.2: Relationship to Functional Framework

Recall that at the end of the previous chapter, a limited, yet functional framework (Figure 2.9.1) was given to illustrate the relationships between each of the theoretical variables discussed in the review of literature. The methodology discussed above aims to contribute to the literature new perspectives on the relationships among these variables and variables not previously investigated. In relation to the framework, the current study aimed to investigate whether another construct exists (specifically, a mathematics-specific psychoeducational facet) and whether this facet is related to the other constructs already illustrated. Specifically, due to the design of the study, the potential links between this facet and inhibitory control, mathematical belief systems, mathematics classroom experiences, mathematical expertise, and learning and developmental experiences, were investigated. Additionally, the study aimed to reexamine the link between inhibitory control and mathematical expertise while also exploring the relationship between inhibitory control and mathematical belief systems.

3.2 Setting and Participants

Before the study was undertaken and before any recruitment had begun, IRB approval was sought and obtained from both the Teachers College, Columbia University IRB as well as the IRB of the participating institution. Once recruitment began, eligible participants for the study were adult (at least eighteen years of age or older), undergraduate students at the same diverse, four-year public university in the northeastern United States. Aligning with Integrated Postsecondary Education Data System (IPEDS) definitions of race and sex, the institution under study is comprised of approximately half white and half non-white students, with nearly two-thirds of the student population being female. Since the study aimed to investigate the potential influence the extent of mathematical knowledge has on problem solving and students' psychoeducational make-up, each participant was actively enrolled in an undergraduate mathematics class at the university. Students who were enrolled in remote (virtual synchronous and asynchronous) classes were not invited to participate since the mathematical and psychoeducational instruments were given on paper.

Furthermore, participants were selected according to the level of mathematics classes in which they were enrolled. Each mathematics course at the university under study requires the satisfaction of a prerequisite requirement. These generally take the form of certain score attained on a placement exam; the transfer of acceptable credits from another postsecondary or secondary institution; or the successful completion (often subject to minimum grade requirements) of a prerequisite course. Therefore, a direct measure of mathematical expertise is required to advance from one course level to another. Indirectly, this provides a leveling structure where courses in successive levels are separated by measurable milestones of mathematical proficiency and expertise. The university defines the academic levels in mathematics in this way according to their academic program structure.

The subgroups were formed after consulting university faculty and the university's course catalog to determine each course's prerequisite requirements; rigor and depth of the mathematical content; and whether these satisfy different subsequent quantitative major requirements. The subgroups of participants were students from introductory (commonly referred to as "remedial" or "developmental") courses in mathematics, which do not satisfy the university requirements in mathematics and serve as prerequisite courses for courses in the next level (e.g., introductory/intermediate algebra); a mathematics course that satisfies the general education requirement for students not in the STEM (science, technology, engineering, or mathematics) majors (e.g., elementary statistics, mathematics for elementary education, etc.); a required mathematics course for STEM majors (e.g., calculus I, calculus II, intermediate statistics, etc.); and a major course in mathematics (e.g., foundations of mathematics, discrete mathematics, real analysis, etc.).

The resulting sample was selected by convenience. Willing faculty members in the mathematics department of the university created opportunities for the research to occur after their regularly-scheduled classes. Then, students from these classes were given the opportunity to voluntarily participate in the research. Data collection occurred between the dates of 4/20/2022 and 5/2/2022, which spanned the thirteenth, fourteenth, and fifteenth weeks of the university's semester. Therefore, students who were not enrolled in mathematics courses at this time were not eligible to participate. The withdrawal deadline for the university under study occurs during the tenth week of classes, and thus, students who withdrew from the class were not eligible to participate.

In total, 130 students elected to participate in the study. Of these 130, 49 (37.69%) were male, 78 (60%) were female, 1 (0.77%) preferred not to provide gender information, and 1 (0.77%) identified as non-binary. Gender information was recorded since several studies in previous literature have noted that gender is an important variable in the analysis of scores on the Cognitive Reflection Test (e.g., Brañas-Garza et al., 2019; Campitelli & Gerrans, 2014; Frederick, 2005). In terms of academic year, 46 (35.38%) students identified as being in their first-year, 26 (20%) were in their sophomore year, 34 (26.15%) were in their junior year, 22 (16.92%) were in their senior year, and 1 identified as not being in any academic year. All of the contacted student participants were undergraduates (this was verified by the faculty member participants), only 1 identified as not being in any academic year. Discussions with faculty members revealed that the reason for this may be that the student is not sure in which academic year they belong due to credit counts, the retaking of certain classes, and non-traditional paths toward degree completion.

In terms of the course leveling structure, students from each course level (developmental, general, STEM, and mathematics) were recruited. In total, 11 (8.46%) were from the developmental level, 55 (42.31%) were from the general level, 44 (33.85%) were from the STEM level, and 20 (15.38%) were from the mathematics level. The courses from which these students were recruited included introductory/intermediate algebra (developmental level); elementary statistics and mathematics for elementary education (general); calculus I, calculus II, and intermediate statistics (STEM); foundations of mathematics, discrete mathematics, and real analysis (mathematics).

3.3 Procedural Overview

As mentioned above, the data were collected from participants in their regularly-scheduled classrooms. In each instance of data collection, willing participants were each administered a packet containing the mathematical problems and the psychoeducational survey. The packets were folded, ledger-sized (11in by 17in) sheets of paper that consisted of four letter-sized sheets. The sheets contained a modified version of the Cognitive Reflection Test (MCRT) and the Short Mathematics Psychoeducational Inventory (SMPI). The former consisted of five mathematics word problems with spaces provided for participants to engage in scratch work, to provide their final answers, and to indicate whether they have seen the problem before. The latter consisted of a 41-item survey questionnaire comprised of 7 demographic and subgroup items as well as 33, 5-point Likert scale items measuring level of agreement to a set of statements, followed by one open-ended space for participants with the prompt “Is there anything else you would like us to know” (for the entire survey, see Appendix B of this document). The development of these instruments and their properties are discussed in detail in the next section. Also in the next section, the procedures, instruments, and analyses are provided according to each research question.

3.4 Procedures for Research Question 1

3.4.1 Instrument

Background Research on the CRT Recall that Research Question 1 states “How do students from different collegiate mathematics courses perform on the Cognitive Reflection Test (modified by the investigator to include two additional problems)?” In pursuit of collecting data to answer this question, a modified version of the Cognitive Reflection Test (CRT) was administered on paper to participating students. The original CRT consists of three mathematical word problems that are “hard to solve, not because of their linguistic or mathematical demands but rather because the correct answer requires the inhibition of a very easy, seemingly correct answer” (Verschaffel et al., 2020, p. 9). Interpreted through the dual-process theories of cognition (Evans & Stanovich, 2013), the CRT measures one’s ability to intervene immediate, intuitive, or default responses to a stimulus (Type 1 processes) with the Type 2 processes that are more reflective, conscious, and which use working memory resources. Moreover, to be successful on these problems, one must activate their inhibitory control capacities and engage in cognitive reflection to effectively inhibit the intuitive, yet faulty response suggested by the problem structure. After successful inhibition of the intuitive response, the solver must then activate their more reflective, Type 2 cognitive processes and perform elementary mathematics operations to determine the correct response.

The CRT was chosen as the instrument to measure inhibitory control in this study due to its demonstrated utility in measuring IC in previous literature, its design as an instrument to measure IC in adults, and its cognitive structure as mathematical problems. As noted in the literature review, the CRT has been characterized as “the dominant measure of adult individual differences in analytic versus intuitive thinking” (Young & Shtulman, 2020, p. 1396) or, in other words, differences in the ability to activate inhibitory control abilities to inhibit Type 1 responses and engage in Type 2 cognitive processes. The CRT was designed to measure such thinking in adults, and has been administered extensively to adult participants, particularly college students since its creation (Brañas-Garza et al., 2019; Frederick, 2005). Despite the fact that several researchers have the critiqued the CRT as an instrument merely measuring numerical and cognitive ability (e.g., Otero et al., 2022), others have defended the CRT as a legitimate measure of inhibitory control or cognitive reflection that cannot be accounted for by numeracy and cognitive ability alone (Liberali et al., 2012; Pennycook et al., 2015; Campitelli & Gerrans, 2014; Cokely & Kelley, 2009; Toplak et al., 2011, 2014). Additionally, unlike many instruments that have been developed to measure IC, the CRT does not rely on physical or spatial properties to produce a situation where inhibition is needed. Van Dooren & Inglis (2015) note that in contrast to such instruments, the CRT is “entirely cognitive: a salient response must be inhibited and some analytical work done before a correct answer can be given” (p. 714).

Each of the problems on the CRT were determined to be accessible to the college students under measure since they do not require an extensive set of prerequisite knowledge to solve. Since all participants have graduated from high school with experience in arithmetic and introductory algebraic concepts, they have the necessary “resources” (from Schoenfeld’s framework) to engage with the problems. Each of these problems can be solved without written calculation; however, the extent of formal calculation required includes the creation and solution of linear equations in a single variable, creation and solution of proportions, and arithmetic operations on fractions. Therefore, the CRT requires all participants to engage with the same elementary mathematical information—this is important since one variable under measure in this study is the potential influence of mathematical expertise on inhibitory control ability. Therefore, the mathematical content of the tasks was made the same for all participants, and at a level of complexity that all participants can engage with. This helps to control for the potential influence that the advanced nature of some participants' mathematical knowledge may have on task performance. The problems of the CRT are given below:

1. A bat and a ball cost $1.10. The bat costs $1.00 more than the ball. How much does the ball cost? ______ cents

2. If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? ______minutes

3. In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake? _______days

(Frederick, 2005, p. 27)

Correct and Incorrect Solutions to the CRT Discussed at length in the literature review section, the first problem of the CRT has been characterized as a "classic dual-process default-interventionist task" (Van Dooren & Inglis, 2015, p. 717). Similar to inconsistency language problems (see Lubin et al., 2013, 2016), the "ball and bat" problem requires the solver to inhibit an intuitive calculative reaction based upon the language of the problem. The intuitive, yet incorrect response to the problem is the answer of 10 cents. Solvers will intuitively, and almost immediately, subtract $\$1.00$ from the $\$1.10$ to arrive at $\$.10$ as an answer, which reverses the "$\$1.00$ more…" part of the problem. This faulty, yet seemingly correct and immediate reaction is precisely that which needs to be inhibited in order to correctly answer the problem. If one engages their IC capacities and can inhibit this immediate reaction (essentially stopping and thinking before responding) they provide themselves the opportunity to engage in active, reflective thought about the task at hand. If this is the case, one can more readily arrive at the correct answer of $\$.05$ (since $\$.05 + \$1.05 = \$.05 + (\$.05 + \$1.00) = \$1.10$).

The second and third problems of the CRT can be viewed as similar to incongruent proportional reasoning tasks (see Gillard et al., 2009b). In the second problem, the solver must inhibit the immediate, intuitive response of 100 minutes. The faulty heuristic that needs to be inhibited is one that leverages proportional reasoning. Commonly, yet incorrectly, solvers will view the proportionality among the "5 machines…5 widgets…5 minutes" and transfer that proportionality to match the structure of the incomplete statement "100 machines…100 widgets…" by responding with "100". The propensity to leverage the intuitive heuristic of proportional reasoning is precisely what needs to be inhibited in this situation. The correct answer is in fact 5 minutes, since the rate of producing widgets by each machine is the same. The problem implies that the production of 1 widget requires 1 machine working for 5 minutes. Therefore, if 100 machines were working for 5 minutes, they would each have produced 1 widget, yielding a total of 100 widgets in all.

The third problem also prompts an intuitive, yet faulty proportional reasoning response. The language of the problem notes that the number of lily pads "doubles" each day, and that the solver is to determine when "half" of the lake is covered. The pairing of the language of these opposite operations in the problem prompts to the solver to immediately apply division by 2 to the number of lily pads, since this would "half" or reverse the doubling of the lily pads each day, yielding an incorrect answer of 24 days. However, this reaction is precisely what needs to be inhibited in this situation. The correct answer is 47 days because the number of lily pads doubles each day, and completely covers the lake in 48 days; therefore, the lake was half-covered on the 47^{th} day. In other words, the lily pads double each day, doubling the lake's lily pads on the 47^{th} day would cover the entire lake.

Notice, that answers for each problem on the CRT can be divided into three categories. The first category is a correct answer. The second is an incorrect intuitive answer. For example, the incorrect intuitive answers for the three problems, respectively, are "$.10" or "10 cents", "100 minutes", and "24 days". Such responses generally indicate that the solver failed to activate their inhibitory control capacities and selected the salient, intuitive, yet faulty response that needed to be inhibited in order to engage in Type 2 thinking. The third category are incorrect and non-intuitive answers given for the three problems. These responses, although incorrect, do not necessarily indicate a failure to activate inhibitory control abilities.

Prior Exposure and Decoy Problems Since the CRT has been extensively administered in previous research, several researchers have noted the influence of prior exposure to the CRT problems due to their popularity (Haigh, 2016; Stieger & Reips, 2016; Toplak et al., 2014). Although some research suggests that the instrument is robust against multiple exposures and that performance on the tasks is stable over time (Bialek & Pennycook, 2017; Stagnaro, 2018), researchers have suggested several steps to be taken to account for this phenomenon (Thomson & Oppenheimer, 2016). One such common suggestion is to simply include a subsequent question after each CRT problem asking the solver if they have seen the problem before (Thomson &

Oppenheimer, 2016). Asking whether participants have seen the problem before provides a way to account for prior exposure. In Thomson & Oppenheimer's (2016) investigation of alternate forms of the CRT, the researchers also included "decoy" problems to "determine the rate of overclaiming prior exposure" (p. 102). They continue to note that the decoy problems "were written to have a similar format of the CRT" (p. 102) and are easy to answer. Both the decoy problems and subsequent questions asking the solver about prior exposure were included in the current study.

Therefore, in the current study, two decoy problems were included in the instrument alongside the original CRT problems, and after each of the problems on the instrument, a question asking the solver whether they have seen each problem before was included. The responses to the decoy problems provide a measure of participants' engagement with the instrument. Since the decoy problems doe not require inhibitory control to solve and are straightforward elementary mathematics word problems, a high success rate on these problems would indicate that participants are engaging intentionally with problems and authentically trying to solve them. This, in contrast to the responses to the actual CRT problems, provides an indication of the difficulty presented by the CRT problems as inhibitory control tasks written to prompt a particular response. Therefore, in the current study, participants' performance on the decoy problems was analyzed in relation to their performance on the CRT problems.

The question following each problem asked the solver "Have you seen this problem before?" Responses to these problems provide both an indication of whether the solver has previously been exposed to the problems, and their propensity to overclaim prior exposure to the problems. In the case of the former, responses to these questions that follow the CRT problems were analyzed in relation to the performance on the CRT problems. Previous research has provided mixed results on whether prior exposure to these problems affects future performance; therefore, these measures will help account for the potential influence of this phenomenon. Descriptive and correlational analyses of responses indicating prior exposure and performance on the CRT problems were conducted. In the case of the latter, high rates of claimed prior exposure to both the CRT problems and the decoy problems would indicate the propensity of participants to overclaim their prior exposure to the CRT problems. Therefore, the responses indicating prior exposure for both the CRT and decoy problems were analyzed using descriptive and correlational analyses.

Problem Ordering and Response Format The ordering of problems on modified forms of the CRT has also been studied. Some research seems to suggest that the earlier the actual CRT problems are ordered in the problem set, the better the performance on those problems; however, other research seems to suggest the opposite (Brañas-Garza et al., 2019). Therefore, in an attempt to not group the CRT questions in one section or another, one decoy question led the problem set, followed by two CRT problems, followed by the second decoy question, and finally followed by the remaining CRT questions.

The problem formats of the CRT have also been studied extensively (Brañas-Garza et al., 2019; Sirota & Juanchich, 2018). Frederick's (2005) original administration of the CRT included just the problems and labeled spaces for the responses (e.g., "___cents", "____minutes"). Other researchers have administered the CRT online, in non-standard orders, and with multiple choice formats (Brañas-Garza et al., 2019; Sirota & Juanchich, 2018). For the current study, the CRT was administered on paper and in the format of Frederick's original study. The only change to note is for the first CRT problem, the "ball and bat problem". In Frederick's study, he provided a labeled space for the response as "___cents". In the current study, a similar space has been included, however it has been labeled with the dollar sign as “$___”. Therefore, the difference lies in the correct response of “5 cents” for Frederick’s and “$.05” for the current study. This was included to further reduce the cognitive demand of the task. Often, those who report a correct response to the problem create a linear equation to solve the problem (e.g., $x + (x + 1) = 1.10$), which results in the answer of $x = .05$. This can be reported directly on the line without having to change the unit to “cents”.

Henceforth, the instrument described above, consisting of the original CRT and Thomson & Oppenheimer’s (2016) decoy questions, given in the aforementioned order, will be referred to as the Modified Cognitive Reflection Test (MCRT). Problems 1 and 4 are the decoy questions, and problems 2, 3, and 5 are the original CRT questions. The MCRT appears in the Appendix of this document as part of the administration packet that was distributed to participants. The problems of the MCRT are given below:

1. A cargo hold of a ship had 500 crates of oranges. At the ship’s first stop, 100 crates were unloaded. At the second stop, 200 more were unloaded. How many crates of oranges were left after the second stop? ____crates

   (Thomson & Oppenheimer, 2016, p. 111)
2. A bat and a ball cost $1.10. The bat costs $1.00 more than the ball. How much does the ball cost? $____

   (Frederick, 2005, p. 27)

3. If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? ____minutes

   (Frederick, 2005, p. 27)

4. An expedition on a mountain climbing trip was traveling with eleven horse packs. Each horse can carry only three packs. How many horses does the expedition need? ____horses

   (Thomson & Oppenheimer, 2016, p. 111)

5. In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake? ____days

   (Frederick, 2005, p. 27)

3.4.2 Data Collection

Previous research has been largely inconsistent on the amount of time that should be allotted to the solver of CRT problems. Frederick's (2005) original study seemed to indicate that the CRT problems were included as part of the large questionnaire that was given to participants containing items concerned with other variables such as risk and time preferences. Other research suggests that approximately two to three minutes should be allotted for a participant to solve all of the CRT problems (Otero et al., 2022). Therefore, since the entire administration procedure for the current study (consisting of the MCRT and the psychoeducational survey) was intended to take no more than twenty-five minutes, participants were encouraged prior to the administration to move on to the survey after ten minutes. However, in an attempt to ensure that time is not a confounding factor for solving the mathematics problems, the participants were permitted to manage their own time within the twenty-five minutes. It is important to note that during data collection, none of the participants took longer than ten minutes on the mathematics problems, and none of the participants needed longer than twenty-five minutes to complete both sections.

Materials, such as pencils and four-function calculators (in an attempt to avoid the possible influence of by-hand computation on MCRT performance), were made available to all participants who needed them. The participants were permitted to use their own calculators (the primary researcher confirmed that the most advanced calculator used was a graphing calculator similar to the Texas Instruments TI-84) and their own writing utensils since pen or pencil were appropriate for the administration packet. The administration packet consisted of several sets of directions for the MCRT (including a set of directions for the problems and for marking answers in the packet), the MCRT, sets of directions for the psychoeducational survey, and the psychoeducational survey. The entire administration packet is given in the Appendix (Appendix A and B) of this document. The administration packet was a folded, ledger-sized (11in by 17 in) that consisted of four letter-sized sheets. After data collection, the packets were scanned using the Snap 11 survey software, after which data validation and additional entry was conducted in the Statistical Package for the Social Sciences (SPSS).

The psychological survey (discussed in detail in a later section) included a set of 7 demographic and subgroup questions prior to 33 Likert scale items and one open-ended question. The 7 demographic and subgroup questions were used in the analysis of MCRT data to identify participant membership to various subgroups. The subgroups that were identified included participants' academic year (freshman/first-year, sophomore, junior, senior, or other with the option to write-in a response), their major (short answer write-in), and whether participants took a mathematics class during their senior year in high school. If participants indicated that they took a mathematics class during their senior year of high school, a subsequent question asked them to specify (as a short answer write-in) which course they took.

3.4.3 Data Analysis

The data were analyzed using descriptive and inferential techniques. All data were analyzed in the IBM Statistical Package for the Social Sciences (SPSS). Descriptive statistics of MCRT performance (number of correct answers provided), performance on each of the individual MCRT problems (including both decoy and CRT problems), performance on the CRT subset of the MCRT, the classification of incorrect answers given on the CRT (intuitive-incorrect or non-intuitive incorrect), and exposure to MCRT problems (i.e., whether participants indicate having seen the problems before) were calculated. These were calculated at the entire study level (all 130 participants) and with respect to each of the subgroups under analysis (i.e., mathematics course level, gender, academic year, etc.). These are displayed in multi-level tables and graphical depictions, and are discussed throughout the results section.

Inferential techniques were also conducted on the MCRT data. These were conducted at the entire study level (all 130 participants) as well as across each of the individual subgroups. The choice of each inferential technique was influenced by the fact that the distributions of individual variables were not normal, individual subgroups differed from one another in terms of sample size, and that the data under analysis was ranked in nature. The distributions of responses on the entire MCRT, and on individual problems of the MCRT, were also not normal. Previous research, such as that of Otero et al. (2022), has noted this phenomenon; particularly, that the CRT does not yield desirable statistical and psychometric properties. Otero et al. (2022) mention that scores on the CRT do not approximate a normal distribution, and that mean scores generally tend toward the minimum and maximum values, causing a “floor” and “ceiling effect” (p. 101614). This was indeed the case with MCRT data in the current study.

Individual subgroups also differed with respect to the number of participants in each group. The largest subgroup contained 55 student participants, and the smallest contained 11. In fact, two of the subgroups contained numbers of participants that were less than 30. Additionally, the variables under analysis were all ranked in nature, which included MCRT performance (scores ranked 0 to 5), CRT performance (scores ranked 0 to 3), performance on individual MCRT problems (two ranks, correct or incorrect), and prior exposure to CRT problems (two ranks, seen before or not seen before). Therefore, when considering the non-normal, ranked data from subgroups of differing, small sample sizes, parametric analysis for the data were not appropriate. That is, the assumptions for parametric analyses such as the Analysis of Variance (ANOVA), Student's T-tests, and Pearson correlation analyses were violated.

Thus, non-parametric inferential techniques were utilized to analyze the MCRT data, which are designed for use in distributions that are inherently not normal. To determine the association between two dichotomous variables, mean square contingency coefficients (phi coefficients, Yule, 1912) were calculated (see IBM, 2013, p. 152, for a description of the SPSS algorithm that was used). The phi coefficient estimates, and can be interpreted in a similar manner to the common parametric Pearson correlation coefficient, where the coefficient takes on values between -1 and 1. That is, decimal values approaching the maximum of 1 indicate a strong, positive association, and decimal values approaching -1 indicate a strong, negative association. Significance values are assigned approximately by strength using SPSS.

To analyze the association between ranked variables related to the MCRT, Spearman Correlation Coefficient (rho) analyses were run due to the ranked nature of the data (Siegel, 1956; Spearman, 1904; Taylor, 1987). The SPSS algorithm for calculating Spearman's rho coefficient used in the current study is based on the work of Siegel (1956) and is described in detail in the SPSS 22 algorithm book (IBM, 2013, p. 641). Much like the Pearson correlation, the

Spearman correlation also measures the strength and direction of the association between two variables. The measure is standardized and can be interpreted in the same fashion as the Pearson correlation, since the coefficient takes on values between -1 and 1. The assumptions for these analyses include that the variables are at least ordinal in nature and that observations under analysis are paired and independent. Also, the analysis assumes that the there is a monotonous relationship between the two variables, since the rho coefficient measures the strength and direction of such a relationship. The assumptions for these analyses in the current study were satisfied. Statistical significance for Spearman analyses is determined through the calculation of a t-test statistic with $n - 2$ degrees of freedom and the determination of a rejection region on the t-distribution. Since the calculation of significance values involves the size of the group under analysis, significance values assigned to coefficients of differing strength can vary by analysis. That is, depending on the size of the group under analysis, a correlation coefficient that is significant in one analysis may not be significant for another analysis if the groups differ in size.

To determine whether two independent groups differ from one another with respect to a variable related to the MCRT, the Mann-Whitney U-Test (also known as the Mann-Whitney-Wilcoxon Test) were run (Ott & Longnecker, 2001; Wilcoxon, 1945; Mann & Whitney, 1947). Assumptions for the Mann-Whitney U-Test include that the dependent variable should be at least ordinal in nature, that the two groups under analysis are independent, and that the observations of the dependent variable are independent. The assumptions for these analyses were satisfied in the current study. The test statistic for this analysis is the U-statistic, which is calculated using the W-statistic from the Wilcoxon rank sum test. Using the test statistic, and the U-distribution (see Dineen & Blakesley, 1973), significance values can be calculated (the details of this calculation, which is based on that of Dineen & Blakesley, 1973, is given in IBM, 2013, p. 655).

To determine if at least one subgroup under analysis (e.g., mathematics course level) differs from another with respect to values for a variable related to the MCRT, a Kruskal-Wallis One Way Analysis of Variance (also known as the Kruskal-Wallis Test, or the Kruskal-Wallis H-Test) was conducted (Kruskal & Wallis, 1952; Ott & Longnecker, 2001; IBM, 2013, p. 661). This is the non-parametric equivalent to the traditional Analysis of Variance (ANOVA) on parametric data. The assumptions for the Kruskal-Wallis test are that the dependent variables is at least ordinal, that the subgroups are independent, and that the observations for the participants in the subgroups are independent as well. These assumptions were satisfied in the current study. Also, since the calculation and interpretation of the test-statistic is based on the χ^2 distribution with $k - 1$ degrees of freedom, it is appropriate to ensure that each subgroup under analysis has a minimum size of 5. The Kruskal-Wallis test-statistic is the H, which can be adjusted for ties in the ranks of the variables. Since the H-statistic is based on χ^2 distribution with $k - 1$ degrees of freedom (with k being the number of subgroups), the result of the Kruskal-Wallis test is often reported using a χ^2 statistic rather than the H itself. This statistic is then used to determine a rejection region on the χ^2 distribution and yield a corresponding significance value.

If the Kruskal-Wallis test determines that a statistically significance difference exists among the subgroups, pairwise analyses were subsequently run to determine exactly which subgroups differ from one another. These post hoc analyses compare each of the subgroups from the Kruskal-Wallis analysis with one another to determine which significant differences exist. The post hoc analyses in the SPSS non-parametric analysis package utilize Dunn's test for multiple comparisons (Dunn, 1961, 1964; IBM, 2013, p. 673) with a Bonferroni correction (Dinno, 2015; Dunn, 1961; Simes, 1986). Dunn's procedure "uses ranks (or successes for the Cochran test) based on considering all samples rather than just the two involved in a given comparison" (IBM, 2015, p. 673). Conducting several pairwise Mann-Whitney U-tests, for example, does not consider all samples all at once. The test-statistic for Dunn's test is a modified z-test statistic that "approximates exact rank-sum test statistics by using the mean rankings of the outcome in each group from the preceding Kruskal-Wallis test" and bases "inference on the differences in mean ranks in each group" (Dinno, 2015, p. 298). Dinno (2015) describes that the act of conducting multiple pairwise comparisons "redefines the meaning of α" (p. 293), which is used to determine statistical significance. Therefore, the Bonferroni adjustment is utilized to account for this phenomenon. In citing Dunn (1961), Dinno (2015) describes that the Bonferroni adjustment "can modify the rejection level for any test by dividing α by the total number of tests and requires a much smaller p-value to reject any test" and that this adjustment "leaves α numerically intact but multiplies the p-value" (p. 293).

3.4.5 List of Analyses

The following is a list of the sets of analyses that were undertaken to answer Research Question 1. These are given in a precise numbering order according to the headings of this document for ease of navigation.

- 4.1.1 Analysis: Entire Sample MCRT Analysis
- 4.1.2 Analysis: Entire Sample MCRT Scores
- 4.1.3 Analysis: Entire Sample Scores on Individual MCRT Problems
- 4.1.4 Analysis: Entire Sample CRT Scores
- 4.1.5 Analysis: Types of Incorrect Answers to MCRT Problems
- 4.1.6 Analysis: Reported Prior Exposure to MCRT Problems
- 4.1.7 Analysis: Mathematics Course Level Subgroup MCRT Analysis
- 4.1.8 Analysis: Descriptive Statistics of MCRT Scores by Mathematics Course Level
- 4.1.9 Analysis: Inferential Statistics of MCRT Scores by Mathematics Course Level
- 4.1.10 Analysis: Additional Subgroup Analyses

- 4.1.11 Analysis: MCRT Scores by Gender
- 4.1.12 Analysis: MCRT Scores by Academic Year
- 4.1.13 Analysis: MCRT Scores by Academic Major
- 4.1.14 Analysis: MCRT Scores by Mathematics Course Taken Senior Year of High School

3.5 Procedures for Research Question 2

3.5.1 Instrument

Content and Creation of the Instrument Recall that Research Question 2 states: "How do students from different collegiate mathematics courses respond to a psychoeducational survey measuring domain-general academic habits of mind, future orientation, and academic self-concept, and mathematics-specific academic habits of mind, beliefs, and past academic experiences?" In pursuit of collecting data to answer this question, a psychoeducational survey instrument was administered to all participants. The instrument comprises items adapted directly from previous research conducted on the same student population currently under measure in investigations of their learning and development (Ben-Avie et al., 2013; Ben-Avie & Darrow, 2019; Darrow, 2016, 2020). Although all the items appearing on the current survey have been administered in previous studies, they have never before been administered as part of the same instrument. As noted in Chapter 2, the research of Ben-Avie and Darrow (2019) and Darrow (2016) investigating general academic habits of mind, future orientation, and academic self-concept was extended by Darrow (2020) to specifically measure these factors within the domain of mathematics. Therefore, the survey items are designed to measure domain-general academic habits of mind, future orientation, and academic self-concept, and mathematics-specific academic habits of mind, beliefs, and past academic experiences.

The survey begins with general demographic questions with multiple choices and options to fill in information that is not presented. Within these are questions that ask the participant to report their academic year; their major; whether the course in which they are enrolled is a prerequisite course or one that satisfies the university's general education requirement; whether they are taking the class for the first time; how they were permitted to enroll in the class (placement or satisfaction of prerequisite); which courses (if any) they took a course in their senior year of high school. Each of these questions were included to identify and analyze subgroups within the sample under measure. The demographic questions from the survey were followed by 33 psychoeducational items that were adapted from previous research. Each of these were coded on the same five-point Likert scale (1=Strongly Disagree, 2=Disagree, 3=Neutral, 4=Agree, 5=Strongly Agree). The final question was "Is there anything else you would like us to know" for which a large open-ended space was given. The full survey for this administration, complete with the demographic questions, the psychoeducational items, and the structure for each is given in Appendix B of this document. This survey will henceforth be referred to as the Short Mathematics Psychoeducational Inventory (SMPI).

The items that were included to measure domain general academic habits of mind, future orientation, and academic self-concept have been adapted from the work of Dr. Michael Ben-Avie and his colleagues (Ben-Avie, 2008, 2013, 2018; Ben-Avie & Darrow, 2018, 2019; Ben-Avie et al., 2012; Ben-Avie & Polka, 2006; Darrow, 2016, 2020; Pang et al., 2016). These items have been administered in previous research to tens of thousands of college students in longitudinal cohort investigations of learning and development (see Ben-Avie et al., 2012 for a complete description of the psychometric properties of these instruments, including reliability and internal consistency analyses). The items that were included to measure mathematics specific academic habits of mind, beliefs, academic self-concept, and past experiences have been adapted from Darrow (2020) and based upon the research of Darrow (2016) and Ben-Avie and Darrow (2019). These items were only previously administered in a pilot study to a small sample of 55 participants in the same mathematics course.

Analysis of the Instrument Although all of the items appearing on the SMPI have been administered in previous research studies, the current study was the first time that they were combined on a single psychometric instrument. That is, the SMPI is the first survey instrument to combine both domain general and mathematics-specific items measuring psychoeducational facets within a single questionnaire. Moreover, despite the fact that a robust theoretical framework exists for the set of domain-general items, and a modest framework for the mathematics-specific items, there is no prior data or analyses that indicate how these items would function together on a single instrument. Therefore, an exploratory factor analysis was conducted to determine the underlying latent structure of the SMPI after its first administration.

Factor analysis is a statistical procedure that aims to determine "whether large sets of variables can be more parsimoniously represented as measures of one or a few underlying constructs" (Fabrigar & Wegener, 2012, p. 17). These underlying constructs (i.e., factors), or "latent variables" are those that can be conceptualized but not directly observed or measured in general, "meaning that we must use some proxy, or set of proxies, in order to gain insight about them" (Finch, 2020, p. 3). Moreover, as Finch (2020) continues to note, "these proxy measures, such as items on psychological scales, are linked to the latent variable in the form of a causal model, whereby the latent variable directly causes manifest outcomes on the observed variables" (p. 3). In the case of the current study, the latent variables are psychoeducational facets, including academic habits of mind, academic self-concept, etc., and the observed variables are the items on the SMPI.

Exploratory factor analysis (EFA) is a type of factor analysis that aims to determine the underlying (or latent) structure of an instrument, identify the latent constructs, and group the measured variables that contribute to each of the latent constructs together (Fabrigar & Wegener, 2012). In contrast, when the particular construct and the items used to measure it are known (and have been developed through previous research), a confirmatory factor analysis (CFA) procedure can be run to determine if sample data is consistent with the measurement structure (Finch, 2020; Long, 1983). CFA analyses can be limiting in that they restrict the analysis model by focusing on a specific set of hypothesized constructs, rather than focusing the investigation on identifying potential latent structures that may exist in addition to what is hypothesized at the onset (Finch, 2020). As mentioned above, it is the case that the items measuring domain general academic habits of mind, future orientation, and academic self-concept included on the SMPI have been developed through previous research. However, they have never been administered alongside items measuring similar traits but specific to the domain of mathematics. Moreover, prior to the administration of the SMPI, there were no data to indicate how these items would interact or whether the inclusion of both of these sets would influence the latent variable structure observed in previous research.

Fabrigar and Wegener (2012) note that there are several assumptions that must be considered when conducting an EFA. One of the assumptions is that the observed measures (in this case, the items on the SMPI) are the “effects” (p. 89) of the latent variables or constructs (in this case psychoeducational facets). This is consistent with the measurement theory that led to the development of the items comprising the SMPI (Ben-Avie & Darrow, 2019; Darrow, 2020). Another assumption is that these effects are linear in nature, which can be violated in cases where the data is ordinal or ranked in nature (Fabrigar & Wegener, 2012), which is the case with the SMPI. Indications of this assumption violation, according to Fabrigar and Wegener (2012), include a poorly fitting model and "implausible or difficult to interpret parameter estimates" (p. 96), which does not appear to be the case with the SMPI.

One assumption of several EFA techniques is that the distribution of variable data are normal in nature (Fabrigar & Wegener, 2012). Since this is not the case with the SMPI, the principal axis factoring method was chosen for the current analysis. Principal axis factoring is a type of factor analysis—which differs from the commonly, and often inappropriately, used method of principal component analysis (see Fabrigar & Wegener, 2012)—that does not require the assumption of normally distributed data (Finch, 2020). Another assumption described by Fabrigar & Wegener (2012) is that a "given measured variable in the battery should not be perfectly accounted for by a linear combination of other observed variables or a linear transformation of another measured variable" (p. 100-101). No such variables exist on the SMPI, and therefore, this assumption is satisfied.

Lastly, there is a great deal of disagreement in the literature regarding the appropriate sample size, number of variables, and the ratio between the number of variables and number of latent constructs for a factor analysis. Several empirical investigations have shown that commonly used rules for these values are inconsistent across studies, depend on other model metrics such as communalities, and that larger sample sizes do not necessarily imply better analyses or a reduction in error (Arrindell & van der Ende, 1985; Costello & Osborne, 2005; de Winter & Dodou, 2012; Hogarty et al., 2005; MacCallum et al., 1999). However, other authors (e.g., MacCallum et al., 1999 references Cattell, 1978, and Lee, 1992) have recommended sizes of at least 200 or greater for a factor analysis. In the current study, there were 130 participants, which falls short of recommended 200 minimum given by many authors, but is above the recommendations of some researchers (e.g., MacCallum et al., 1999 references Gorsuch, 1983, and Kline, 1979). The ratio between observed variables (31) to the number of participants (130) is greater than 1:4, which is considered low, but still within the standards of some previous research (McCallum et al., 1999 references Cattell, 1978). However, it is noted here that the sample size is not desirably large and the ratio between the number of variables and the sample size is also not desirable in considering the recommendations of many authors, which may influence the results. Therefore, this fact will be used to approach any findings with caution and contextualize any results.

Due to the nature of the data and the satisfied assumptions given above, a principal axis factoring analysis was conducted on the data. In general, EFA analyses inherently produce an infinite number of solutions rather than a single unique solution, since there are “an infinite number of combinations of the factor loadings that will yield the same mathematical fit to the data” (Finch, 2020, p. 39). This is a procedural consequence of the analysis and is remedied through the use of a rotational method. Finch (2020) continues to explain that a “rotation refers to a mathematical transformation of the initially extracted set of factor loadings with the goal of relating each observed indicator variable to a single latent factor” (p. 39). Therefore, through the use of a rotation, the latent factors and the individual variables related to these factors can be identified and interpreted.

There are two classes of rotations, orthogonal rotations and oblique rotations. Orthogonal rotations require the assumption that the factors are uncorrelated to one another (Fabrigar & Wegener, 2012). In the current study, this assumption is violated since the items of the SMPI measure both domain general and mathematics-specific psychoeducational facets, these are likely to be correlated since they both are related in academic situations. Oblique rotations, in contrast, do not require the assumptions that factors be uncorrelated. Therefore, an oblique rotational method was chosen for the current analysis. Arguably "the most commonly used of all rotations, both oblique and orthogonal, is Promax (Hendrickson & White, 1964)" (Finch, 2012, p. 48), which was used in the current study. The Promax rotation, Fabrigar & Wegener (2012) note, begins with an orthogonal rotation (specifically, the Varimax rotation after a Kaiser normalization, see IBM, 2013, p. 292-293) and "then conducts a mathematical transformation of this initial solution, by raising factor loadings to a power of two or greater, to arrive at an oblique solution" (p. 78). The power to which the loadings are raised is known as kappa, and is most commonly used value for this is 4. Finch (2020) notes that "in practice, it is unusual for a researcher to alter kappa from the default value of 4" (p. 48). Therefore, consistent with previous research and accepted convention for psychological research of the type used in the current study, a Promax rotation with kappa equal to 4 was used.

The factor analysis was conducted in IBM SPSS (information regarding these algorithms can be found in IBM, 2013, p. 330-345). The output of the analysis includes a correlation matrix (depicting the correlations between each of the items in the analysis and the others); tests regarding the suitability of the data for factor analysis; information about the eigenvalues (which represent "the variances of latent variables obtained through a mathematical decomposition of the correlation matrix" Finch, 2020, p. 19) and percentage of total variance explained by the factors; a Scree plot (which plots and traces ordered pairs consisting of the factor number and its corresponding eigenvalue); the pattern matrix (which contain the loadings in the form of regression coefficients); the structure matrix (which contain the correlations between individual items and the observed factors); and the factor correlation matrix (which depicts the correlations between each of the factors).

After the correlation matrix of the output was analyzed, the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy (Kaiser, 1970) was calculated (see IBM, 2013, p. 344) and analyzed. Essentially, the measure provides “an assessment of whether the variables belong together psychometrically and thus whether the correlation matrix is appropriate for factor analysis” (Dziuban & Shirkey, 1974, p. 359). According to Kaiser & Rice (1974), values above .8 for this measure are preferrable, and to use their words, those above .9 are “marvelous”, those above .8 are “meritorious”, and those above .7 as “middling” (p. 112). In practice, values nearing .8 or exceeding .8 are generally perceived as acceptable values, and that those lower than .7 are “mediocre” and those lower than .5 are “unacceptable” (Kaiser & Rice, 1974, p. 112). Another test, Bartlett’s Test of Sphericity (IBM, 2013, p. 485) was also conducted to determine whether the correlation matrix calculated in a factor analysis is significantly different from the identify matrix. The test utilizes a Chi Square test statistic to determine a significance value. Taken together, the KMO measure and Bartlett’s test of sphericity indicate whether the data is suitable for a factor analysis.

An analysis of the eigenvalues and the total variance explained by the model was undertaken. It is general practice to retain factors with eigenvalues greater than 1 (this is commonly referred to as the Kaiser criterion). However, many researchers have advised against using this as the sole criterion for retaining factors (Fabrigar & Wegener, 2012; Gorsuch, 1980). In addition, the Scree plot (Catell, 1966) should be analyzed, which plots the eigenvalues against the individual factors. The points are connected, and one looks for the “last major drop” in these connections to indicate the number of factors that are defined by the model (Fabrigar & Wegener, 2012, p. 57). Additional analyses that can be run at this stage include parallel analysis (which was explored in the current analysis, see Chapter 5 for discussion of this and directions for future analyses).

The output of the analysis also includes a pattern matrix and a structure matrix. The pattern matrix, Fabrigar & Wegener (2012) note, contains the actual factor loadings and "are comparable to standardized partial regression coefficients" (p. 80). The structure matrix contains the correlations between individual items and the factors identified by the model. There exists some disagreement on how to appropriately interpret the factors using these two matrices. Generally, the literature suggest that researchers should use both matrices to interpret the nature of each of the factors (Gorsuch, 1983). However, Fabrigar and Wegener (2012) note that "the pattern matrix should be the primary basis for interpreting factors" (p. 81), since this follows "the logic of oblique rotations" (p. 81). In general, the selection of the appropriate number of factors, as Fabrigar and Wegener (2012) note, is "best addressed in a holistic fashion" and such a decision "is as much theoretical as it is statistical" (p. 55). Therefore, in the current analysis, the pattern matrix was indeed used as a primary basis for interpretation, but the structure matrix was also used to strengthen this interpretation and contribute to the construction of reliable scales.

Lastly, the factor correlation matrix indicates the correlations between the individual factors. The correlations in this matrix for the current analysis were anticipated to be moderate, since the constructs under measure are each related to participants' academic profile. The Promax rotational method was chosen for this reason. However, this matrix provides important information about how the factors interact with one another on the SMPI after its first administration, which aids in the refinement of the instrument for future administrations (see Chapter 5 for a discussion of future research directions).

Now, based upon the information presented above, the results of the factor analysis conducted on SMPI data in the current study will be detailed. Based upon the conditions discussed above, a principal axis factoring procedure with a Promax rotation was used. The KMO measure was calculated and Bartlett's Test of Sphericity was also conducted. Recall that these were all coded on the same 5-point Likert scale (1=strongly disagree, 2=disagree, 3=neutral, 4=agree, 5=strongly agree) measuring level of agreement with the given statement. The KMO measure was .768, which is an acceptable value (just below the "meritorious" level) and the Bartlett's Test of Sphericity was significant [$\chi^2(528) = 1616.861, p < .001$], which together indicate that the data was suitable for factor analysis. Prior to extraction, there were 10 factors that emerged with initial eigenvalues greater than 1. However, only the first five of these had a well-defined and interpretable structure. That is, the first five factors each contained groups of itmes that could be interpreted as measuring a common construct, or that could be directly interpreted through the theoretical framework utilized to justify their inclusion on the instrument. The other factors that were not well-defined seemed to group items that were not related to one another or whose connection did not have a meaningful interpretation in context.

The summary measures for the first 5 factors are given below in Table 3.5.1, which include the initial eigenvalues, percent of total variance explained by each factor, and the cumulative percent of variance explained by the factors. It is important to note that these metrics are related to total variance, rather than common variance, which is provided after factor extraction and retention. In the table below, and the succeeding tables, the factors are labeled according to the items that were included in each and the underlying theory (from previous research, e.g., Ben-Avie & Darrow, 2019 and Darrow, 2020) that contributed to their inclusion on the SMPI. The labeling of these factors is described in detail in the discussion of the pattern and structure matrices below. The factors are: Limiting Beliefs, Habits, and Experiences Related to Mathematics (LBHEM); Academic Habits of Mind (AHM); Future Orientation (FO); Academic Self-Concept (ASC); and Hesitancy to Engage with Faculty (HEF).

Table 3.5.1

Factor Metrics for Principal Axis Factor Analysis on 33-Item SMPI

Factor	Initial Eigenvalues	Percent of Total Explained Variance (Prior to Extraction)	Percentage of Cumulative Explained Variance (Prior to Extraction)
LBHEM	5.968	18.083	18.083
AHM	4.924	14.922	33.005
FO	2.045	6.197	39.202
ASC	1.906	5.775	44.977
HEF	1.470	4.456	49.433

Note. Limiting Beliefs, Habits, and Experiences Related to Mathematics (LBHEM); Academic Habits of Mind (AHM); Future Orientation (FO); Academic Self-Concept (ASC); and Hesitancy to Engage with Faculty (HEF).

Analysis of the Scree plot also indicated that the first four factors should be retained, with the fifth coming just after the “last major drop”. Thus, the Scree plot indicates the existence of a four-factor model, however, the fifth factor was analyzed further due to its structure and its straightforward interpretability in context. Figure 3.5.2 displays the Scree plot for the analysis.

The pattern and structure matrices for the analysis are given in Tables 3.5.3 and 3.5.4, respectively. Recall that the pattern matrix contains the actual loadings of individual items on each factor. The structure matrix provides the correlations between individual items and the factors. In these tables below, the rows are the individual items of the SMPI, and the columns indicate each factor given with the labeling described above.

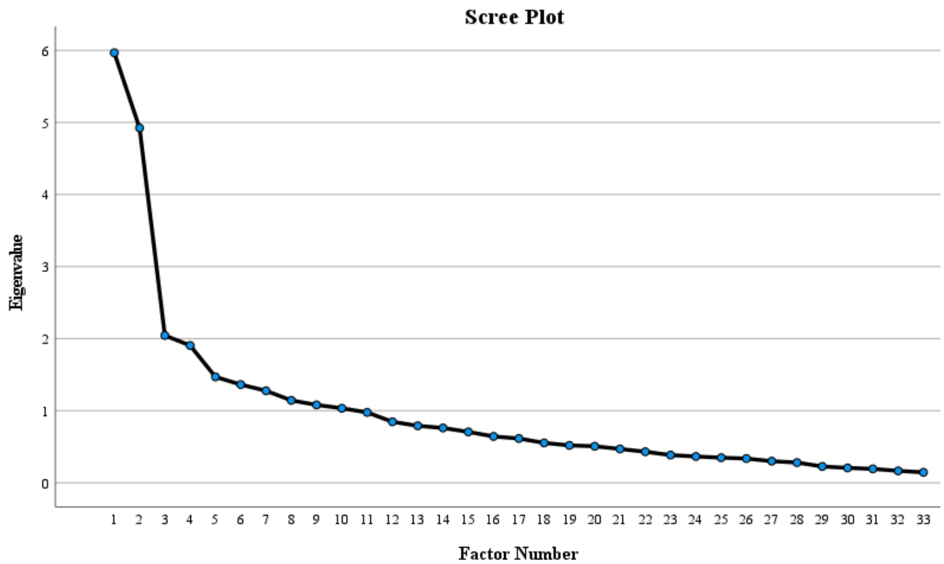

Figure 3.5.2: Scree Plot for Principal Axis Factor Analysis on 33-Item SMPI

For the ease of viewing and factor identification, loadings (in the pattern matrix) and correlations (in the structure matrix) below .35 were suppressed. In analyzing the pattern matrix, the structure of the first factor is well-defined, and does not contain any items with cross loadings on other factors (note again that only values above .35 are being shown). The structure of the factor is also straightforward in terms of its interpretation, since the items that belong to the first factor deal with self-limiting beliefs, academic habits of mind, and past educational experiences related to mathematics. The factor is structured in a way such that it relates limiting characteristics, or those that are associated with less desirable outcomes. That is, items for which higher levels of agreement correspond to less desirable outcomes are positively related to the factor (i.e., the actual factor loadings detailed in the pattern matrix are positive). For example, the first item of the factor, which also had the strongest loading, was one measuring perhaps the most general and pervasive self-limiting belief related to mathematics "I'm just not good at math". The loading for this item was strong and positive, indicating that higher levels of agreement to this

Table 3.5.3

Pattern Matrix for Principal Axis Factor Analysis with Promax Rotation on 33 SMPI Items

SMPI Item	Factors				
	LBHEM	AHM	FO	ASC	HEF
I'm just not good at math.	.998				
Math and/or anything with numbers has been an obstacle to my academic success.	.857				
I have experienced difficulties in math since high school or before high school.	.800				
I am usually confident that I will do well on math tests.	-.670				
I can explain how I reach the correct answer on a math test.	-.572				
I push aside math assignments and do them last.	.528				
The wording of math problems confuses me.	.391				
I often play catch-up in my classes.		-.696			
I wait until right before a math test to start studying.		-.633			
I give myself enough time to really read course materials.		.438			
I am definitely a "work before play" type of person.		.364	.364		
I settle for just passing my courses.		-.353			
I break down long-term assignments and/or class projects and work on them over time.		.352			
I am taking the necessary steps to obtain the career I desire.			.930		
I have a fairly clear idea of what I need to study now in order to have the career I want.			.694		
Thinking about the future I want makes me do more now to get that future.			.444		
My confidence in academic skills has increased this semester.				.805	
I am doing better than I thought I would in college.				.711	
I fear that if I ask for help, my professor will think less of me.					.839
I am hesitant to raise my hand in class even though I know the answer.					.656
I tell my professor when I don't understand something from math class.					-.565

Note. Only the first five factors are detailed here. Also, only loadings greater than .35 are shown

Table 3.5.4
Structure Matrix for Principal Axis Factor Analysis with Promax Rotation on 33 SMPI Items

SMPI Item	Factors				
	LBHEM	AHM	FO	ASC	HEF
I'm just not good at math.	.875				
Math and/or anything with numbers has been an obstacle to my academic success.	.817				
I am usually confident that I will do well on math tests.	-.767			.421	
I have experienced difficulties in math since high school or before high school.	.690				
The wording of math problems confuses me.	.636			-.363	
I can explain how I reach the correct answer on a math test.	-.629			.388	
I push aside math assignments and do them last.	.515				
I expect to use the math I have learned in my future career.	-.515				
I wait until right before a math test to start studying.		-.651			
I often play catch-up in my classes.		-.612		-.358	
I break down long-term assignments and/or class projects and work on them over time.		.569	.399	.442	
I give myself enough time to really read course materials.		.545			
I am definitely a "work before play" type of person.		.508	.451		
I am taking the necessary steps to obtain the career I desire.			.763		
Thinking about the future I want makes me do more now to get that future.		.355	.653	.429	
I have a fairly clear idea of what I need to study now in order to have the career I want.			.533		
I study regularly to be successful in college.		.506	.509		
My confidence in academic skills has increased this semester.				.670	
I am doing better than I thought I would in college.				.599	
I am always well-prepared for math class.	-.410			.499	
I settle for just passing my courses.			-.406	-.446	
When something confuses me, I think about it until I can make sense out of it.				.355	
I fear that if I ask for help, my professor will think less of me.					.774
I am hesitant to raise my hand in class even though I know the answer.					.763
I tell my professor when I don't understand something from math class.					-.593
I use my time between classes productively.		.457	.411	.409	
I find it hard to prioritize my time.		-.452		-.432	
Sometimes I am disappointed in my test results because I studied a great deal.	.414	.459			

item are positively related to the factor, and thus indicating that less desirable outcomes and beliefs are characterized by the factor. Next, were the two items measuring self-limiting beliefs in mathematics developed through sentiments about past academic experiences in the discipline, "Math and/or anything with numbers has been an obstacle to my academic success," and "I have experienced difficulties in math since high school or before high school." Responses to these items indicate the nature of one's academic self-concept and belief systems in mathematics. Like the first item of the factor, higher levels of agreement to these items are associated with less desirable self-perceptions and belief systems, particularly with respect to describing participants' academic history in mathematics.

The next item of the factor for which the loading was negative was "I am usually confident that I will do well on math tests", which indicates that responses indicating a lack of confidence with mathematics assessments. This is also the case with the items "I can explain how I reach the correct answer on a math test" and "I expect to use the math I have learned in my future career". The former indicates an academic habit of mind related to mathematical work related to study skills, attention to detail, reflection, and confidence. The latter indicates a belief about learning mathematics and perception of its utility.

The remaining items of the factor also articulate sentiments related to academic habits of mind. Although these were positively related to the factor, they indicated negative outcomes in mathematical academic habits of mind, which aligns with the factor structure of characterizing less desirable or limiting attributes. These items were "I push aside math assignments and do them last" and "The wording of math problems confuses me". The former indicates a habit of mind related to study habits and independent work as well as how mathematical work is valued and prioritized. The latter indicates a habit of mind related to the engagement with mathematical content and also reflects a belief about the nature of mathematics. In viewing the structure matrix, other items that articulated limiting academic habits of mind related to mathematics were correlated with the factor. These were, “I am always well-prepared for math class” and “Sometimes, I am disappointed in my test results because I studied a great deal”. Therefore, when taken together, the items of this factor collectively describe a set of beliefs, academic habits of mind, and self-conceptions based upon past academic experiences that are limiting in the endeavor of learning mathematics. Thus, consistent with the structure of the factor, it has been named Limiting Beliefs, Habits, and Experiences Related to Mathematics (LBHEM). This name will also be sued to describe the scale that was developed from the factor (discussed in detail in the next section).

The next factor was also well defined, despite two items that were cross-loaded onto two different factors (this is discussed in the succeeding paragraphs). The structure of the factor is also straightforward in terms of its interpretation, since the items that belong to the second factor deal with general academic habits of mind, related to study skills, time management, work habits, and academic mindset. In contrast to the previous factor, the structure of this factor indicates that more desirable outcomes in these areas are positively related to the factor. For example, the first item on the factor, was “I often play catch-up in my classes”. This item had the strongest loading on the factor, but it was negative, indicating that higher levels of agreement to this item are associated with less desirable outcomes in general. Therefore, responses (lower levels of agreement) are positively related to the factor. Such responses indicate an academic diligence and desirable levels of study skills, time management, and quality independent work.

This is also the case with the next item, “I wait until right before a math test to start studying”, where the loading is also negative. Thus, lower levels of agreement to this item (indicating more desirable study habits, i.e., not waiting until the last minute to begin studying) are positively related to the factor. This is also true for the item "I settle for just passing my courses". This item indicates academic mindset and habits of mind related to coursework. Higher levels of agreement to this item are associated with a complacency with lower course grades, less academic motivation, and a lesser commitment to academic work. The final two items of the factor, "I give myself enough time to really read course materials" and "I am definitely a 'work before play' type of person", are positively related to the factor since higher levels of agreement are associated with more desirable outcomes in time management and study skills. In viewing the structure matrix, other items that articulated habits of mind related to time management and study skills were correlated with this factor. These included "I study regularly to be successful in college," "I use my time between classes productively," and "I find it hard to prioritize my time." In alignment with the structure of the factor characterizing desirable academic habits of mind, it has been named Academic Habits of Mind (AHM).

It should be noted that the item "I am definitely a 'work before play' type of person" is significantly cross-loaded with the third factor (which will be described below as a factor measuring future orientation). In fact, the two loadings are identical. Such a phenomenon generally indicates a flaw in the questionnaire or with the data on these items. However, the cross-loading is reasonable when interpreted within the context of both of the factors on which it loaded. On the one hand, this item articulates an academic habit of mind consistent with desirable study skills. Those who "work before play" prioritize their work as more important than recreation, which is a desirable characteristic within the academic domain. On the other, those who do this have a mature future time perspective; they are able to understand that prioritizing their work is more important in the long-term than recreation. Thus, these individuals are oriented to the future, and thus, have desirable levels of future orientation, which is what the items comprising the third factor characterize. Although this interpretation is consistent with the structure of the factors and the theoretical foundation informing the creation of the SMPI, it is unusual for both the item to have the exact same loading on both factors. Since this was the case, this item was omitted from the construction of reliable psychometric scales (described in detail below) using SMPI items.

As mentioned above, the third factor was well-defined and related to future orientation. The three items in this factor each had positive loadings, indicating higher levels of agreement of those items were positively associated with the factor. The first item of the factor, which had the highest loading, was “I am taking the necessary steps to obtain the career I desire”. This item articulates the goal-directed action that is essential to future orientation. Responses to this item indicate the nature of one’s future time perspective, and the ability conceive of, and take actions in the here-and-now to influence their future. The next item is “I have a fairly clear idea of what I need to study now in order to have the career I want”, which also indicates he ability to conceive of one’s own development and identify goal-directed actions. The third and final item of the factor (with the exception of the item described in the previous paragraph that was cross-loaded with the AHM factor) is “Thinking about the future I want makes me do more now to get that future”, which also indicates the ability to conceive one’s own development and future, but also provides information about how such a conception motivates one to take the goal-direction actions in the present.

Items from the structure matrix that were correlated with the factor articulate mature study habits and time management. Although these are more aligned with the structure and nature of the AHM factor, these can be interpreted as being related to future orientation as well.

For example, the item “I study regularly to be successful in college” indicates study skills and goal-directed behavior. Responses to this item indicate one’s perception of the utility of studying regularly and that doing so supports the future goal of being successful in college. However, since these items were in general more strongly related to the AHM factor and more closely aligned with the structure and nature of the AHM factor, they were not pursued further as potential items for a future orientation scale. That is, these items contributed to AHM scale development but not to future orientation scale development (described further in a succeeding section). Consistent with its structure, the factor has been named Future Orientation (FO).

The fourth factor is only comprised of two items, and is less comprehensive in scope than the others. The two items in this factor are related to one’s confidence in their academic skills and their academic self-concept. The loadings of both items in the pattern matrix were strong and positive. Since higher levels of agreement on both items indicated more desirable levels of confidence and academic self-concept, the factor characterizes positive academic self-concept. The first item, “My confidence in academic skills has increased this semester” had the strongest of the two loadings. This item indicates growth in academic skills and growth in positive academic self-concept. The second item is “I am doing better than I thought I would in college” which, in contrast to the self-limiting beliefs described in other factors, reflects positive academic self-concept and a growth in self-perception. Together, these two items provide an indication of one’s view of themselves within the academic domain of college. They also indicate growth in this respect; the first item indicates growth in academic skills and in confidence, and the second indicates growth in one’s view of themselves as a result of self-reflection. Thus, since this factor comprises items indicating one’s academic self-concept, it has been accordingly named Academic Self-Concept (ASC).

The fifth and final factor identified in this analysis comprises three items related to students' interactions and engagement with faculty members. The first item of the factor, which also has the strongest loading, is "I fear that if I ask for help, my professor will think less of me". This item indicates a self-limiting belief that has the potential to impact learning, since a hesitancy to seek help when needed could have detrimental academic effects. The loading for this item was positive, suggesting that higher levels of agreement are positively related to the factor. This was also true of the second item, "I am hesitant to raise my hand in class even though I know the answer", which articulates a self-limiting behavior that has the potential to reduce positive learning effects in the classroom. Thus, just in viewing the first two items, this factor seems to characterize self-limiting beliefs and behaviors related to engaging with faculty. The last item of the factor, "I tell my professor when I don't understand something from math class", indicates a healthy classroom behavior that is related to positive learning outcomes. This item was negatively associated with the factor, which suggests that responses associated with a hesitancy to engage with faculty were positively related to the factor. It should be noted that there was a small typo printed on the survey involving this item, however the nature of this was unlikely to influence how participants responded to the item.

Taken together, these items indicate self-limiting behaviors and beliefs that describe a hesitancy to engage with faculty members even in a time of need. Thus, consistent with the structure, the fifth and final factor has been named Hesitancy to Engage with Faculty (HEF). In viewing the structure matrix, a number of other items on the SMPI were correlated with this factor, and many of these were related to the AHM factor. This is reasonable since these items measured academic habits of mind. Those with desirable levels of academic habits of mind are also likely to have a positive view of themselves in the academic domain as a result. However, these items are more closely aligned with, and were intended to measure academic habits of mind. Therefore, these were not included in the development of scales using the HEF factor and instead contributed to scales developed from the AHM factor.

Each of the factors described above were clearly defined by the actual factor loadings given by the pattern matrix, and were interpretable in context as latent constructs describing psychoeducational facets. Supporting evidence gleaned from the structure matrix also aided in characterizing each factor according to the items to which it was related. The factors themselves were also related to one another, which was anticipated at the onset of the study. The correlations between each of the factors are given below in Table 3.5.5. The rows and the columns are indicated by the factor labels, and the cells are the correlation coefficients. To reduce the amount of redundant information presented, only the lower half of the matrix is given, which provides each combination of bivariate correlations between the five identified factors.

Table 3.5.5

Factor Correlation Matrix for Principal Axis Factor Analysis with Promax Rotation on 33 SMPI Items

Factors	LBHEM	AHM	FO	ASC	HEF
LBHEM	1.00	-	-	-	
AHM	.156	1.00	-	-	
FO	-.089	.392	1.00	-	-
ASC	-.324	.320	.475	1.00	-
HEF	.174	.086	-.086	-.142	1.00

Prior to the analysis, it was anticipated that any factors including mathematics specific forms of items measuring self-limiting beliefs and academic habits of mind would be correlated with those including domain general forms of these items. This was the reason for the choice of the Promax rotational method. Indeed, this was the case, where the correlation between LBHEM and AHM was .156. The correlation between Limiting Beliefs, Habits, and Experiences Related to Mathematics (LBHEM) and Academic Self-Concept (ASC) was -.324, indicating a moderate, negative relationship, suggesting higher levels of limiting characteristics related to mathematics are associated with lower levels of academic self-concept. Also consistent with the structure of the factors is the moderate, positive relationship (.320) between academic habits of mind and academic self-concept. This suggests that more desirable academic habits are positively associated with positive academic self-concept. The largest correlation to note in the matrix was that between academic self-concept and future orientation and (.475). This suggests that higher levels of academic self-concept are related to higher levels of future orientation. This is reasonable since both are related to more desirable academic outcomes. Additionally, responses to the items comprising the academic self-concept factor indicate growth and positive self-concept as a result of that growth. This could provide someone with evidence that they are able to grow and achieve their hoped for future as a result.

Scale and Composite Score Development Part of the aim of the factor analysis and the administration of the SMPI was to develop reliable scales to measure psychoeducational facets related to academic work both generally and with respect to mathematics. Based upon the analysis, it seems as if the SMPI is a multidimensional instrument measuring latent constructs related to these facets. To determine the reliability of each scale in the current administration, reliability analyses utilizing the reliability coefficient Cronbach's Alpha (α) were used (this is also known as tau-equivalent reliability). In citing earlier work (Kelly, 1942), Cronbach (1951) notes that "a reliability coefficient demonstrates whether the test designer was correct in expecting a certain collection of times to yield interpretable statements about individual differences" (p. 297). That is, the instrument (often referred to as a test in literature) has a certain degree of internal consistency. Tavakol & Dennick (2011) explain that "internal consistency describes the extent to which all the items in a test measure the same concept or construct and hence it is connected to the inter-relatedness of the items within the test" (p. 53).

In citing Komperda et al. (2018), Barbera et al. (2021) note that Cronbach's Alpha should be described as a "measure of single-administration reliability" (p. 257) rather than a measure that implies consistency in administrations beyond the one under measure. One assumption of a Cronbach Alpha analysis for a scale is that the scale is designed to measure a single construct. In citing Miller (1995), Tavakol and Dennick (2011) write that "fundamentally, the concept of reliability assumes that unidimensionality exists in a sample of test items and if this assumption is violated it does cause a major underestimate of reliability" (p. 54). In the factor analysis described above, five well-defined factors emerged that seem to describe several underlying psychoeducational facets. The items that comprise these factors seem to characterize the facets themselves. Some of the factors measure clear unidimensional psychoeducational facets that have been described in previous research, such as academic self-concept, academic habits of mind, future orientation, and hesitancy to engage with faculty.

Although the factor measuring limiting beliefs, habits, and experiences is more complicated to define, it still seems to measure self-perceived limitation related to mathematics. Future research is certainly needed to refine the SMPI and the scales that comprise it to better measure and characterize exactly the latent psychoeducational facet described by the LBHEM factor (see Chapter 5 for a discussion of future directions). However, at the current time, there seems to be evidence that the LBHEM factor is measuring something particular and potentially unidimensional. Therefore, reliability analyses were still run on the scale developed from this factor with the knowledge that the results may be underestimates of the scale's true reliability and that refinements to the scale will be undertaken in future research.

Another assumption to be considered in such a reliability analysis is related to the measure of Cronbach's Alpha. Tavakol and Dennick (2011) describe that Cronbach's Alpha "is grounded in the 'tau equivalent model' which assumes that each test item measures the same latent trait on the same scale" (p. 54). All of the items within each individual scale are coded on the same 5-point Likert scale. However, for some items, lower values of the scale (1=strongly disagree, 2=agree) may algin with higher values on another item (4=agree, 5=strongly agree). For example, in the LBHEM factor, higher values for the item "I'm just not good at math" are associated with lower values for the item "I am usually confident I will do well on math tests" which each indicate less desirable or limiting outcomes. Therefore, in the development of the scales described below, items needed to be reverse coded so that each item aligns with one another on the same measured 5-point scale.

The measure of Cronbach's Alpha is standardized and varies between 0 and 1, where values approaching 1 indicate greater reliability. In citing Taber (2018), Barbera et al. (2021) note, as many researchers have in recent years, that "there is no standard, threshold or criterion value for an acceptable alpha" (p. 258). However, Tavakol & Dennick (2018) note that a common convention in research involving psychometric instruments is that values between $\alpha =$ .7 and $\alpha = .95$ are considered acceptable (Tavakol & Dennick, 2018, cite Nunnally, 1994, Bland, 1997, DeVellis, 2003, as reporting on this convention). Other writers have noted that coefficients above $\alpha = .6$ are also commonly accepted (e.g., Hair et al., 2006). Each of the scales included on the SMPI were subject to a Cronbach Alpha analysis, and are included below. For each of the following, the name of the scale matches the factor from which it was derived.

The first scale was developed from the LBHEM factor. The items in this scale were taken directly from the pattern matrix (Table 3.5.3) describing the actual loadings for the items in each factor. Although a number of other items from the SMPI were moderately correlated with the LBHEM factor, the items from the pattern matrix formed the core of the scale, and reliability metrics were not improved with the addition of other variables from the structure matrix (Table 3.5.4). Since the LBHEM factor measures limiting characteristics related to mathematics, less desirable responses on each item are positively associated with the factor. Therefore, two items needed to be reverse coded to match the structure of the factor so that reliability analyses could be run. These items were "I am usually confident that I will do well on math tests" and "I can explain how I reach the correct answer on a math test". In total, there are 7 items on the scale developed from the LBHEM factor, and the scale shares the same name as the factor from which it was derived. A tau-equivalent reliability analysis was conducted using Cronbach's Alpha on this scale (see IBM, 2013, p. 821, for the SPSS algorithm used). The analysis revealed that $\alpha = .860$, which suggests that the scale is strongly reliable since the Cronbach Alpha measure is well above $\alpha = .7$ and is close to $\alpha = .9$. The items of the scale are given below along with indications of which items were reverse coded.

Scale: Limiting Beliefs, Habits, and Experiences Related to Mathematics (LBHEM Scale). (Reliability: $\alpha = .860$)

- I'm just not good at math.
- Math and/or anything with numbers has been an obstacle to my academic success.
- I have experienced difficulties in math since high school or before high school.
- I am usually confident that I will do well on math tests. (Reverse coded)
- I can explain how I reach the correct answer on a math test. (Reverse coded)
- I push aside math assignments and do them last.
- The wording of math problems confuses me.

The next scale was developed from the AHM factor. Each of the items appearing in the pattern matrix for this factor were included, with the exception of the item “I am definitely a ‘work before play’ type of person.” This item was omitted because it had the exact same loading (.364) for the AHM and FO factor. Although it is reasonable to expect this item to be related to both academic habits of mind and future orientation, it was omitted from scale development for this reason. Additionally, several items from the structure matrix that were moderately correlated with the AHM factor were also included in the scale. The item “I find it hard to prioritize my time” was included and was moderately correlated with the overall factor (.457). Although the item was correlated with the factor of ASC, it was more strongly correlated with the AHM factor and articulates a sentiment related to academic habits. Also included from the structure matrix was the item “I study regularly to be successful in college”. This item was moderately correlated with both the AHM factor (.506) and the FO factor (.509), with similar correlation coefficients. It is reasonable to expect that this item would be moderately correlated with both factors. On the one hand, it articulates an academic habit of mind related to study skills. On the other, it articulates a future time perspective and goal-directed behavior. However, the language of the item aligns more closely with the language articulated in the other items of the AHM factor and provides a domain-general indicator of study skills, complimenting the item “I wait until right before a math test to start studying” which could be interpreted as an academic habit specific to mathematics. Moreover, in the reliability analyses on the AHM scale and FO scale, the inclusion of the item on the AHM scale substantially increased the reliability score of the scale (by .39), whereas its inclusion on the FO scale resulted in very little increase (by .005). Therefore, for these reasons, the item was included on the AHM scale and excluded from the FO scale. The decision not to include the item on both scales also reduces the repeated measure of variance on different scales presumably measuring different psychoeducational constructs.

In total, there are 7 items on the scale developed from the AHM factor (5 taken directly from the pattern matrix, and 2 from the structure matrix), and the scale shares the same name as the factor from which it was derived. A tau-equivalent reliability analysis was conducted using Cronbach's Alpha on this. The analysis revealed that $\alpha = .742$, which suggests that the scale is reliable since the Cronbach Alpha measure is above $\alpha = .7$. The items of the scale are given below along with indications of which items were reverse coded.

Academic Habits of Mind (AHM Scale)
(Reliability: $\alpha = .742$)

- I often play catch-up in my classes. (Reverse coded)
- I wait until right before a math test to start studying. (Reverse coded)
- I give myself enough time to really read course materials.
- I settle for just passing my courses. (Reverse coded)
- I break down long-term assignments and work on them over time.
- I study regularly to be successful in college.
- I find it hard to prioritize my time. (Reverse coded)

The next scale was developed from the FO factor. The items in this scale were taken directly from the pattern matrix. Although a number of other items from the SMPI were moderately correlated with the LBHEM factor, the items from the pattern matrix formed the core of the scale, and reliability metrics were not improved with the addition of other variables from the structure matrix. Moreover, as described above, such items from the structure matrix were included on other scales where the relationship between the two was observed to be stronger. In total, there are 3 items on the scale developed from the FO factor, each taken directly from the pattern matrix. As with the other scales, the scale shares the same name as the factor from which it was derived. A tau-equivalent reliability analysis was conducted using Cronbach's Alpha on this. The analysis revealed that $\alpha = .637$, which is lower than the conventional $\alpha = .7$ threshold but is still considered acceptable. The items of the scale are given below along with indications of which items were reverse coded.

Future Orientation (FO Scale)
(Reliability: $\alpha = .637$)

- I am taking the necessary steps to obtain the career I desire.
- I have a fairly clear idea of what I need to study now in order to have the career I want.
- Thinking about the future I want makes me do more now to get that future.

The next scale was developed from the ASC factor. The items in this scale were also taken directly from the pattern matrix. However, there were only two items in the factor and on the scale. Generally, it is recommended that psychometric scales consist of at least three items (Marsh et al., 1998, as cited in Robinson, 2017). Having at least three items is important for effective parameter estimation, comprehensiveness of the scale, and bias reduction. Therefore, it is noted here that the scale developed from the ASC factor does not meet standard recommendations regarding minimum number of items and is limited by this (see Chapter 5 for directions for future research, including item development and testing regarding the ASC factor). Moreover, results that relate to the scale developed from the ASC factor will be approached with caution and the number of items on the scale will help contextualize any findings.

The limitations of the scale also limit the interpretability of a tau-equivalent reliability analysis was conducted using Cronbach's Alpha . With this in mind, the result of this analysis on the two ASC items revealed that $\alpha = .635$, which is lower than the conventional $\alpha = .7$ threshold, but is still considered acceptable. As with the other scales, the scale shares the same name as the factor from which it was derived. The two items of the scale are given below, both of which are positively coded.

Academic Self-Concept (ASC Scale)
(Reliability: $\alpha = .635$)

- My confidence in academic skills has increased this semester.
- I am doing better than I thought I would in college.

The final scale was developed from the HEF factor. The items in this scale were taken directly from the pattern matrix. Unlike the other factors, there were no other items in the structure matrix that were moderately correlated with the HEF factor. Only the three core items with actual factor loadings were those related to the factor. Therefore, this factor is unique in that the construct being measured seems to be relatively isolated and unrelated to the other factors. This is also observable in the factor correlation matrix (Table 3.5.5), where all the correlations between the HEF factor and others are low and are the lowest in the table overall. As with the other scales, the scale shares the same name as the factor from which it was derived. A tau-equivalent reliability analysis was conducted using Cronbach's Alpha on this. The analysis revealed that $\alpha = .733$, which suggests that the scale is reliable since the Cronbach Alpha measure is above $\alpha = .7$. The items of the scale are given below along with indications of which items were reverse coded.

Hesitancy to Engage with Faculty (HEF Scale)
(Reliability: $\alpha = .733$)

- I fear that if I ask for help, my professor will think less of me.
- I am hesitant to raise my hand in class even though I know the answer.
- I tell my professor when I don't understand something from math class. (Reverse coded)

Using the scales that were described and analyzed above, composite scores were created for the purposes of analyzing responses related to the psychoeducational facets described by each factor/scale. The composites were created by directly adding the responses for the items on each scale. Recall that each of the scales were coded in the same structure (some items needed to be reverse coded before inclusion on the scale). The composite score for the LBHEM scale, referred to as LBHEM Composite, consisted of 7 items, when directly added resulted in a minimum composite score of 8 and a maximum of 35. The mean of the LBHEM Composite was 20.09, the median was 20, and the standard deviation was 6.2.

The AHM Composite was calculated by summing the 7 AHM Scale items, and ranged from 9 to 34, with a mean of 22.93, a median of 22, and a standard deviation of 4.9. The FO Composite was calculated by summing the 3 FO Scale items, and ranged from 5 to 15, with a mean of 11.99, a median of 12, and a standard deviation of 1.84. The ASC Composite was calculated by summing the 2 ASC Scale items, and ranged from 2 to 19, with a mean of 6.67, a median of 7, and a standard deviation of 1.844. Lastly, the HEF Composite was calculated by summing the 3 HEF Scale items and ranged from 3 to 15, with a mean of 8.07, a median of 8, and a standard deviation of 3.153. Each composite was also standardized (utilizing Z-Scores) for use in regression analyses.

3.5.2 Data Collection

As mentioned in the procedures for Research Question 1, the SMPI was administered to all 130 participants in the same session as the MCRT was administered. All 130 participants finished the MCRT before moving on to the SMPI. All 130 participants responded to most of the items on the SMPI, including the demographic questions at the beginning of the survey. Each participant finished within the allotted twenty-five minutes given for both the MCRT and SMPI.

It should be noted that the SMPI was taken after the MCRT was taken. Since several of the items on the MCRT deal with mathematics beliefs, habits of mind, and past experiences, the participants may have been primed for these questions after completing the MCRT. That is, participants' responses to items regarding mathematics may have been slightly influenced by the experience of taking the MCRT right before taking the SMPI (see Chapter 5 for a discussion of this potential limitation and directions for future administrations of the SMPI).

3.5.3 Data Analysis

The data were analyzed using descriptive and inferential techniques. All data were analyzed in the IBM Statistical Package for the Social Sciences (SPSS). Descriptive statistics of responses to individual items of the SMPI were calculated for the entire sample and for each individual subgroup under analysis. The subgroups were identified through the responses to the demographic questions appearing at the beginning of the survey. The subgroups that were identified included participants' academic year (freshman/first-year, sophomore, junior, senior, or other with the option to write-in a response), their major (short answer write-in), and whether participants took a mathematics class during their senior year in high school. If participants indicated that they took a mathematics class during their senior year of high school, a subsequent question asked them to specify (as a short answer write-in) which course they took.

Inferential techniques were also conducted on the SMPI data. These were conducted at the entire study level (all 130 participants) as well as across each of the individual subgroups. The choice of each inferential technique was influenced by the fact that the distributions of individual variables were not normal, individual subgroups differed from one another in terms of sample size, and that the data under analysis was ranked in nature (5-point Likert scales). The distributions of responses to items on the SMPI were not normal. Therefore, when considering the non-normal, ranked data from subgroups of differing, small sample sizes, parametric analysis for the data are not appropriate. The assumptions for parametric analyses such as the Analysis of Variance (ANOVA), Student's T-tests, and Pearson correlation analyses were violated.

Thus, non-parametric inferential techniques were utilized to analyze the SMPI data, which are designed for use in distributions that are inherently not normal. To determine the association between individual variables, and between composite scores, Spearman Correlation Coefficient (rho) analyses were run due to the ranked nature of the data (Siegel, 1956; Spearman, 1904; Taylor, 1987). As mentioned before, the algorithm used for the analysis is described in detail in the SPSS 22 algorithm book (IBM, 2013, p. 641). The nature of this analysis as well as the assumptions that are required were discussed in the procedures for Research Question 1. The assumptions for a Spearman correlational analysis run on SMPI data were satisfied.

To determine whether two independent subgroups differ from one another with respect to responses to an item of the SMPI, Mann-Whitney U-Tests (also known as the Mann-Whitney-Wilcoxon Test) were run (Ott & Longnecker, 2001; Wilcoxon, 1945; Mann & Whitney, 1947). The nature of this analysis as well as the assumptions that are required were discussed in the procedures for Research Question 1. The assumptions were satisfied for Mann-Whitney U-Tests run on SMPI data (the details of this calculation are given in IBM, 2013, p. 655).

To determine if at least one subgroup under analysis (e.g., mathematics course level, academic major, etc.) differs from another with respect to responses to an item on the SMPI, a Kruskal-Wallis One Way Analysis of Variance (also known as the Kruskal-Wallis Test, or the Kruskal-Wallis H-Test) were conducted (Kruskal & Wallis, 1952; Ott & Longnecker, 2001; IBM, 2013, p. 661). This is the non-parametric equivalent to the traditional Analysis of Variance (ANOVA) on parametric data. The nature of this analysis as well as the assumptions that are required were discussed in the procedures for Research Question 1. The assumptions were satisfied for Kruskal-Wallis tests run on SMPI data. To determine, which subgroups differed from the others, six pairwise post hoc analyses were run. The post hoc analyses in the SPSS non-parametric analysis package utilize Dunn's test for multiple comparisons (Dunn, 1961, 1964; IBM, 2013, p. 673) with a Bonferroni correction (Dinno, 2015; Dunn, 1961; Simes, 1986).

3.5.4 List of Analyses

The following is a list of the sets of analyses that were undertaken to answer Research Question 2. These are given in a precise numbering order for ease of navigation according to the section number.

- 4.2.1 Analysis: SMPI Item Analysis
- 4.2.2 Analysis: Descriptive Statistics of SMPI Items Across Entire Sample
- 4.2.3 Analysis: SMPI Item Analysis Across Mathematics Course Level Subgroups
- 4.2.4 Analysis: SMPI Item Analysis Across Other Subgroups
- 4.2.5 Analysis: SMPI Scale Analysis
- 4.2.6 Analysis: SMPI Scale Analysis Across Mathematics Course Level Subgroups
- 4.2.7 Analysis: SMPI Scale Analysis Across Other Subgroups

3.6 Procedures for Research Question 3

3.6.1 Instrument

Recall that Research Question 3 states "What is the relationship between the performance of students from different collegiate mathematics courses on a modified version of the Cognitive Reflection Test and their responses to a psychoeducational survey measuring domain-general academic habits of mind, future orientation, and academic self-concept, and mathematics-specific beliefs, habits of mind, and past academic experiences?" In pursuit of answering this question, responses to both the MCRT and SMPI were analyzed in relation to one another.

Therefore, the instruments under analysis are the two described in detail in the previous sections. Data collected from the MCRT and data collected from the SMPI were directly compared.

3.6.2 Data Collection

As mentioned in the procedures for Research Question 1 and Research Question 2, the MCRT and the SMPI was administered to all 130 participants in the same session. All 130 participants finished the MCRT before moving on to the SMPI. All 130 participants responded to most of the items on the SMPI, including the demographic questions at the beginning of the survey. Each participant finished within the allotted twenty-five minutes given for both the MCRT and SMPI. It should be again noted that the SMPI was taken after the MCRT was taken.

3.6.3 Data Analysis

The data were analyzed using descriptive and inferential techniques. All data were analyzed in the IBM Statistical Package for the Social Sciences (SPSS). To analyze the strength and direction of the relationship between MCRT scores and responses to individual SMPI items and composite scores (developed through the scales of the SMPI), Spearman correlational analyses were run due to the ranked nature of the data of the MCRT scores, individual SMPI items, and composite scores. As mentioned before, the algorithm used for the analysis is described in detail in the SPSS 22 algorithm book (IBM, 2013, p. 641). The details of this procedure and the required assumptions, which are satisfied for the analyses involving MCRT and SMPI data, are given in the procedures for Research Question 1. Spearman correlational analyses were run between the aforementioned measures for each of the subgroups under analysis.

To determine the predictive power of composite scores and subgroup membership on MCRT performance, multivariate binary logistic regression analyses were run. Logistic regression analyses determine the influence of a set of independent variables on a dichotomous outcome variable (Borden, 2005). Membership to one of the dichotomous outcomes is predicted by the independent variables. Then, the prediction is compared based upon the actual membership of individuals one of the dichotomous outcomes. Within the actual analysis, the outcome variable is transformed into the log odds ratio or logit. Borden (2005) describes that the logit is "the natural log of the quantity: the probability of belonging to one group divided by the probability of belonging to the other group" (p. 144). In the current study, the probability P of belonging to the zero score group is divided by the probability of belonging to the non-zero score group $(1 - P)$. Then, the natural logarithm of this quantity, $\ln\left(\frac{P}{1-P}\right)$, is the logit, represented as $logit(P)$.

The logistic regression analysis yields an equation, where the independent variables are the inputs, whose coefficients are calculated in a way that minimizes error in prediction. These coefficients "represent the unit change in the outcome (in this case logit(P)) for each unit change in the predictor" (Borden, 2005, p. 144). However, these coefficients predict a logarithmic quantity, and therefore, when they are exponentiated, they can be directly interpreted. In particular, Borden (2005) continues to describe that these exponentiated coefficients "reveal how many times more likely it is to obtain the outcome…for each unit change in the predictor" (p. 144).

To determine whether or not the model is statistically significant (that is, significantly different from the null model, or the model with no variables) a Chi Square (χ^2) test statistic is calculated with degrees of freedom $k - 1$, with k being the number of predictor variables in the model. The Hosmer-Lemeshow goodness-of-fit test is also calculated to determine if the model suitably fits the data (Hosmer & Lemeshow, 2000; IBM, 2013, p. 536; Peng et al., 2002). If the result of the test is not statistically significant, then the model suitably fits the data. The strength of prediction and utility of the logistic regression analysis can be assessed using the classification summary and the metrics of the model, including measures such as the Cox & Snell R^2 and the Nagelkerke R^2 value. The final classification table provides a summary of how well the model predicted group membership to the dichotomous groups indicated by the outcome variable. This is calculated by comparing the actual group membership from the observed data and the predicted membership provided by the model. The measures of the Cox & Snell R^2 (Cox & Snell, 1989) and the Nagelkerke R^2 (Nagelkerke, 1991) are known as Pseudo-R^2 values. In multiple regression analyses, the adjusted R^2 (adjusted for the presence and influence of multiple predictor variables) is a value that, in essence, measures the explained variance in the outcome variable by the predictor variables. The Snell R^2 and the Nagelkerke R^2 were designed to provide a similar metric for logistic regression analyses. However, these measures are highly variable and depend on the individual model, and therefore, they should be "interpreted with great caution" (Borden, 2005, p. 145).

There are several assumptions for a logistic regression analysis. First, the outcome variable must be binary in nature. In the current study, this is the case, where responses to the CRT subset of the MCRT were separated into two groups: zero score and non-zero score. The proportion of all participants who earned a zero score on the CRT was substantial, with 87 (66.9%) of participants not answering any of the CRT questions correctly. Therefore, a binary grouping of participants with zero score and non-zero scores was formed. Thus, the first assumption is satisfied. Second, the observations of the predictor variables must be independent. Although the MCRT and SMPI data were collected from individuals independently, the responses of the SMPI came in time after the MCRT. In the current analysis, the SMPI data is being used to predict MCRT data, but the MCRT was taken before the SMPI. Therefore, as mentioned before, a priming effect may be present between the two sets of data. Therefore, it is noted that the independence assumption is potentially violated, and the results will be approached with caution and will be interpreted with this fact in context.

A third assumption is that the predictors are not strongly correlated with one another. In each logistic regression analysis in the current study, predictors were not combined in the same model if they were correlated to one another. A fourth assumption assumes there are not substantial outliers. An analysis of the SMPI data (including composite scale scores) and MCRT data (which are ranked scores of performance from 0 to 3) showed no influential outliers since the responses were bounded on similar ranked scales.

A fifth assumption is that there is a linear relationship between the logit values and the continuous predictor values for each predictor. Graphical analysis of model results revealed that this was the case for the composite score predictor of the binary outcome variable. There was only one model with a continuous predictor variable, since the others involved only "dummy variables" (dichotomous variables indicating subgroup membership). These are permitted as predictors in logistic regression, but do not form a linear relationship with the logit variable. Lastly, there is an assumption of sufficient sample size in each of the outcome groups, with a "commonly recommended minimum" of 10 to 20 observations in each independent variable (Stoltzfus, 2011, p. 1). Crosstabulations of values in each group of the dummy variables and in each composite score with the outcome variable of zero and non-zero scores on the CRT confirm this assumption is satisfied.

3.6.4 List of Analyses

The following is a list of the sets of analyses that were undertaken to answer Research Question 3. These are given in a precise numbering order according to the section headings of this document for ease of navigation.

- 4.3.1 Analysis: Correlational Analyses of MCRT and SMPI Items
- 4.3.2 Analysis: Correlational Analysis of MCRT Scores and SMPI Responses Across Entire Sample
- 4.3.3 Analysis: Correlational Analysis of MCRT Scores and SMPI Responses Across Mathematics Course Level Subgroups
- 4.3.4 Analysis: Correlational Analysis of MCRT Scores and SMPI Responses Across Other Subgroups
- 4.3.5 Analysis: Correlational Analyses of MCRT and Composite Scores
- 4.3.6 Analysis: Correlational Analysis of MCRT and Composite Scores Across Entire Sample
- 4.3.7 Analysis: Correlational Analysis of MCRT and Composite Scores Across Mathematics Course Level Subgroups
- 4.3.8 Analysis: Correlational Analysis of MCRT and Composite Scores Across Other Subgroups
- 4.3.9 Analysis: Logistic Regression Analyses
- 4.3.10 Analysis: Logistic Regression Analyzing Mathematics Course Level Subgroup Membership in Predicting Zero or Non-Zero CRT Scores
- 4.3.11 Analysis: Logistic Regression Analyzing LBHEM Scale Composite Score in Predicting Zero or Non-Zero CRT Scores

Chapter 4: Results

The results of the current study are presented here according to each research question. The analyses conducted to answer each research question are given in a numbered format.

4.1 Research Question 1: How do students from different collegiate mathematics courses perform on a modified version of the Cognitive Reflection Test (modified by the investigator to include two additional problems)?

4.1.1 Analysis: Entire Sample MCRT Analysis

The following group of analyses focus on performance on the MCRT by all participants. For reference, the MCRT is included again below:

1. A cargo hold of a ship had 500 crates of oranges. At the ship's first stop, 100 crates were unloaded. At the second stop, 200 more were unloaded. How many crates of oranges were left after the second stop? ____crates

   (Thomson & Oppenheimer, 2016, p. 111)

2. A bat and a ball cost $1.10. The bat costs $1.00 more than the ball. How much does the ball cost? $____

   (Frederick, 2005, p. 27)

3. If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? ____minutes

   (Frederick, 2005, p. 27)

4. An expedition on a mountain climbing trip was traveling with eleven horse packs. Each horse can carry only three packs. How many horses does the expedition need? ____horses

   (Thomson & Oppenheimer, 2016, p. 111)

5. In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake? ____days

   (Frederick, 2005, p. 27)

Recall that the decoy problems are problems 1 and 4, and that the problems from the original CRT are problems 2, 3, and 5. Also recall that answers for each problem on the CRT can be divided into three categories. The first category is a correct answer. The second is an incorrect intuitive answer. For example, the incorrect intuitive answers for the three problems, respectively, are "$.10" or "10 cents", "100 minutes", and "24 days". Such responses generally indicate that the solver failed to activate their inhibitory control capacities and selected the salient, intuitive, yet faulty response that needed to be inhibited in order to engage in Type 2 thinking. The third category are incorrect and non-intuitive answers given for the three problems. These responses, although incorrect, do not necessarily indicate a failure to activate inhibitory control abilities.

4.1.2 Analysis: Entire Sample MCRT Scores

The following graph (Figure 4.1.1) depicts the breakdown of performance of the entire sample on the MCRT according to the number of correct answers provided by the participants. Each bar of the graph represents the number of individuals earning a particular score (1 MCRT problem correct, 2 MCRT problems correct, etc.) and the number of individuals in each of these groups is given below each bar. The vertical axis indicates the relative frequency percentage of the number of individuals in each group out of the total number of participants (130). The data table used to form this graph is given in Appendix C (Table 4.1.1).

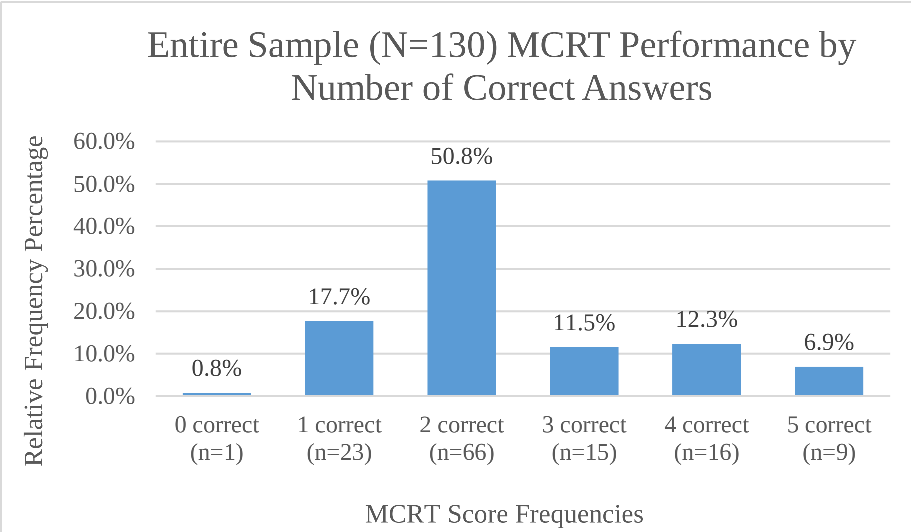

Figure 4.1.1: Graphical Depiction (Bar Graph) of Entire Sample MCRT Performance by Number of Correct Answers

The majority (81.53%) of all participants answered at least 2 of the MCRT problems correctly. The mean of all MCRT scores was 2.3769 with a standard deviation of 1.143. The median of all correct responses was whereas the median was 2 as was the interquartile range. Recall that the MCRT was comprised of two decoy problems (1 and 4) and the three original CRT problems (2, 3, and 5). The results on the MCRT therefore, are influenced by the fact that the two decoy problems were easier to solve since they did not involve an intuitive conflict like the CRT problems. In fact, of the 66 individuals who scored a 2 on the MCRT, 64 (97%) of them answered problem 1 correctly and all 66 (100%) of them answered problem 4 correctly. This implies that 97% of individuals who scored a 2 on the MCRT answered only the decoy questions correctly, and incorrectly answered each of the CRT problems. Additionally, there were 40 participants who scored at least a 3 on the MCRT and nearly all of these participants answered each of the decoy problems correctly. This means that all except two participants who correctly responded to any of the CRT problems correctly necessarily provided correct answers to each of the decoy problems.

4.1.3 Analysis: Entire Sample Scores on Individual MCRT Problems

Performance on each individual problem on the MCRT is depicted in the graph (Figure 4.1.2) below. There are two sets of bars in the graph for each problem indicating performance. The green bars indicate the percentage of correct answers given to each problem (out of the total of 130 participants) and the red bars indicate the percentage of incorrect answers given to each problem (out of the total of 130 participants). The data table used to make this graph is given in Appendix C (Table 4.1.2).

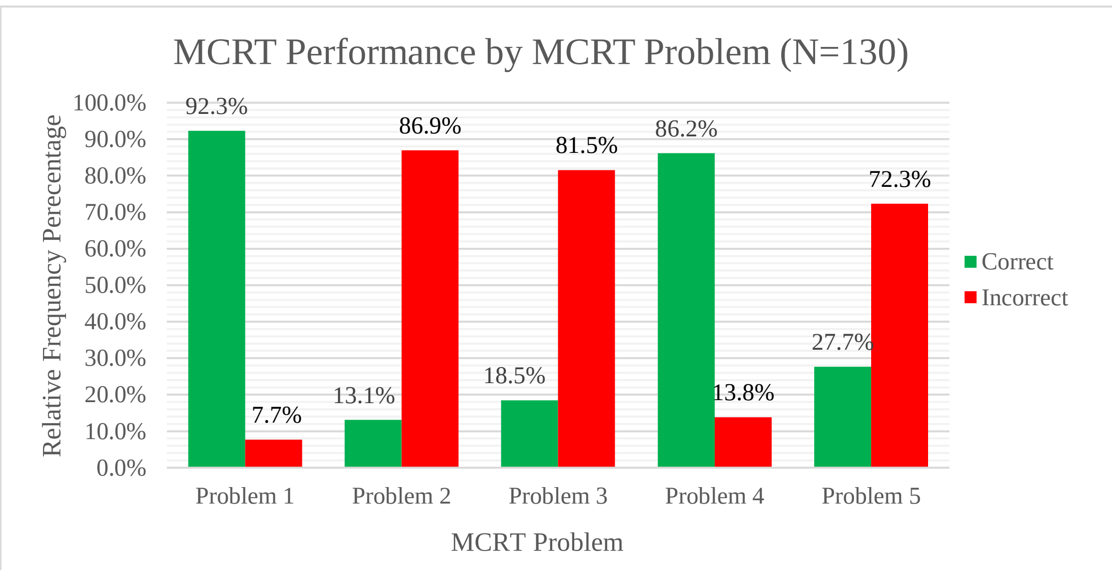

Figure 4.1.2: Graphical Depiction (Side-By-Side Bar Graph) of Entire Sample MCRT Performance by MCRT Problem

As mentioned above, the majority of participants solved both of the decoy problems correctly. Recall that both decoy problems differed from the CRT problems in that they did not present an intuitive interference and were substantially easier to solve as a result. Therefore, one interpretation of the large proportion of correct answers given on the decoy problems provides an indication that participants were solving each of the problems intentionally. For if the participants were not engaged with the task of the MCRT and were not intentionally trying to solve each problem correctly, the proportion of incorrect answers for the decoy problems would be higher and the proportion of incorrect answers given for the decoy and the original CRT problems would be more similar.

For each of the CRT problems, the majority of all participants did not provide the correct answer. This coupled with the success rate on the decoy problems indicates that the nature of the CRT problems influenced participants' responses. It appears that the second problem (the "ball and bat" problem) was the most difficult to solve in general, since 86.9% of participants responded incorrectly, followed closely by 81.5% on the third problem (the "widgets" problem), and finally by 72.3% on the fifth problem (the "lily pads" problem).

4.1.4 Analysis: Entire Sample CRT Scores

The three-item CRT, which is a subset of the MCRT, provides information about participants' ability to activate inhibitory control. The graph (Figure 4.1.3) below provides the scores of the entire sample on the CRT according to the number of correct answers earned on the CRT problems. The bars of the graph indicate the performance of participants according to the number of correct answers provided to each of the CRT problems. The vertical axis provides the relative frequency percentage of these performances out of the total 130 participants. The data table used to make this graph is given in Appendix C (Table 4.1.3).

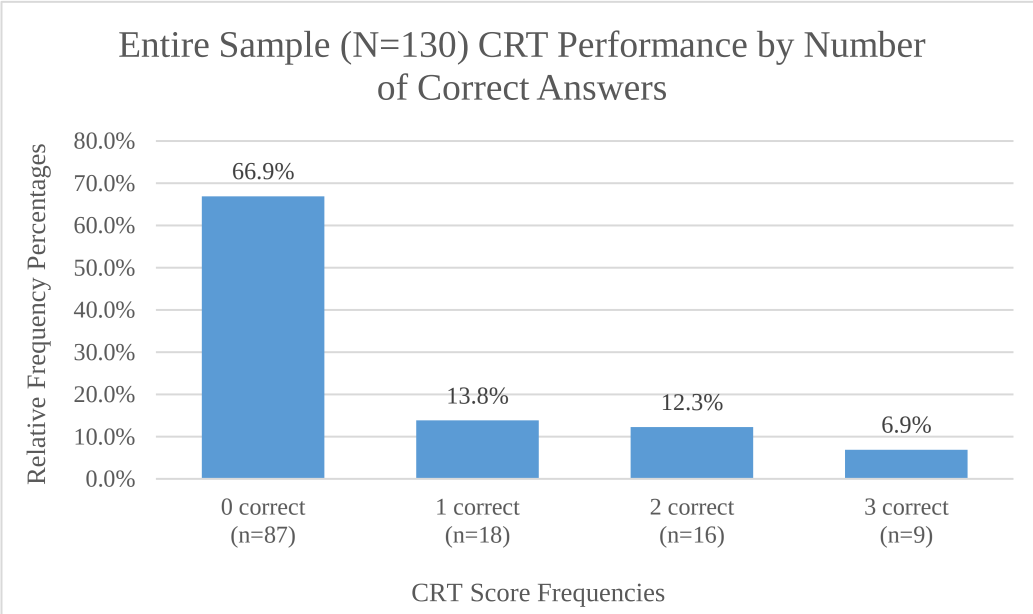

Figure 4.1.3: Graphical Depiction (Bar Graph) of Entire Sample CRT Performance by Number of Correct Answers

The majority of all participants (66.9%) did not correctly solve any of the CRT problems. Additionally, only 33.1% solved at least one CRT problem correctly. Of the 87 participants who did not answer any CRT questions correctly, 80 (92%) answered the first decoy question of the MCRT correctly, and 70 (80.5%) answered the second decoy question correctly. This suggests that the nature of the CRT problems as default-interventionist tasks influenced participants' ability to be successful with the problems. Moreover, the high success rate on the decoy problems but low success rate on the CRT problems indicates that the majority of participants who scored a 0 on the CRT problems were intentionally engaging with the MCRT. Furthermore, this indicates that it was the nature of the CRT problems as inhibitory control tasks that led to participants' incorrect answers, and suggests that the ability to activate inhibitory control on these problems may be underdeveloped in the population under study.

Of the 43 participants who answered at least one CRT problem correctly, 40 (93%) of them answered the first decoy question correctly and 42 (97.7%) answered the second decoy question correctly. Additionally, of those who answered at least one CRT problem correctly, only 17 (39.5%) solved the first CRT problem ("ball and bat") correctly, 24 (55.8%) solved the second CRT problem correctly ("widgets"), and 36 (83.7%) correctly solved the last CRT problem ("lily pads"). This suggests that of those who had a non-zero score on the CRT found most success with the third CRT problem, followed by the second CRT problem, and finding most difficulty with the first CRT problem. This is consistent with the entire sample analysis that suggested the first CRT problem ("ball and bat") was the most difficult for participants overall.

Of the 25 participants that scored at least a 2 on the CRT, all 25 (100%) answered both decoy questions correctly. Additionally, of these 25 participants, 16 (64%) answered the "ball and bat" problem correctly, 20 (85%) answered the "widgets" problem correctly, and 23 (92%) answered the final "lily pads" question correctly. This seems to follow the hierarchy of difficulty for the CRT problems observed above with those who scored at least 1 CRT problem correctly, with the "ball and bat" problem being most difficult, followed by the "widgets" problem, and finally by the "lily pads" problem.

4.1.5 Analysis: Types of Incorrect Answers to MCRT Problems

Unlike the CRT problems of the MCRT, the decoy problems do not have multiple classifications of incorrect answers. The decoy problems do not prompt an intuitive response through the language and structure of the problem unlike the CRT problems. For the first decoy problem (problem 1), the majority of all participants (120, 92.3%) provided the correct answer of 200 crates. All of the remaining 10 (7.7%) participants provided the incorrect answer of 300 crates. For the second decoy problem (problem 4), the majority of all participants (112, 86.2%) provided the correct answer of 4 horses. The next most commonly provided incorrect answer of

33 horses was given by 10 (7.7%) participants.

In contrast to the decoy problems, there are two classifications of incorrect answers for the CRT problems since they are default-interventionist tasks requiring inhibitory control. The two types of incorrect responses are intuitive-incorrect responses and non-intuitive-incorrect responses. For the former, these responses indicate that the solver failed to activate inhibitory control capacities, and chose the intuitive response prompted by the language and structure of the problem. For example, the intuitive-incorrect response to the "ball and bat" problem is '$.10', for the "widgets" problem it is '100 minutes', and for the "lily pads" problem it is '24 days'. The following graph (Figure 4.1.4) depicts the breakdown of incorrect responses for each of the CRT problems. The segmented bar graph provides the percentage (out of a total 100%) of each type of incorrect answer classification given to each CRT problem of the MCRT. The table used to form this graph is given in Appendix C (Table 4.1.4).

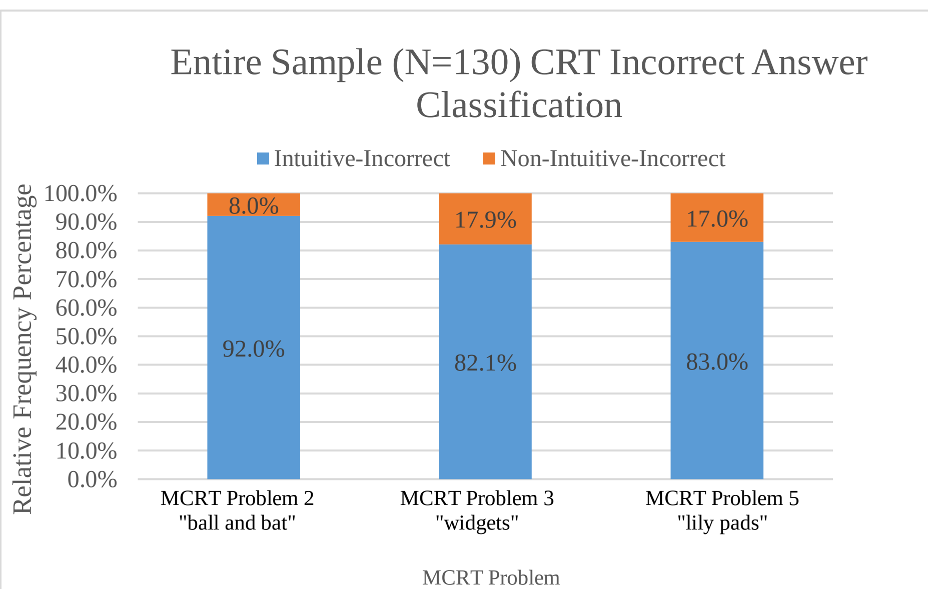

Figure 4.1.4: Graph (Segmented Bar Graph) Depicting Entire Sample CRT Incorrect Answer Classification

For the first CRT problem ("ball and bat"), out of the 113 participants who incorrectly solved the problem, 104 (92%) provided the intuitive-incorrect answer. This pattern was also emerged for the other two CRT problems. Out of the 106 participants who incorrectly solved the "widgets" problem, 87 (82.1%) provided the incorrect-intuitive response; and for 94 participants who incorrectly solved the "lily pads" problem, 78 (83%) provided the intuitive-incorrect response. This suggests that the structure and language of the CRT problems influenced participants' performance since these problems specifically prompt the intuitive-incorrect response. Moreover, these results suggest that the participants who responded incorrectly to the CRT problems failed to effectively activate their inhibitory control capacities, leading to their incorrect responses.

4.1.6 Analysis: Reported Prior Exposure to MCRT Problems

The potential influence of prior exposure to problems on the CRT has been noted in previous research (Thomson & Oppenheimer, 2016). To account for this phenomenon, a follow up problem was asked after each MCRT problem that stated, "Have you seen this problem before?". One aim of including the decoy problems, which were adapted directly from Thomson & Oppenheimer's (2016) study, was to determine participants' propensity to overestimate their exposure to the CRT problems. The following table (Table 4.1.5) provides the breakdown for the responses to the follow up question about prior exposure. For each MCRT problem given in the rows of the table, the columns provide the number (n) and relative frequency percentage (%) calculated out of all 130 participants, respectively, of those who reported seeing the problem before, not seeing the problem before, and those who did not respond to the follow up question about prior exposure.

Table 4.1.5

Entire Sample CRT Prior Exposure to MCRT Problems

MCRT Problem	Seen Before		Not Seen Before		Did Not Respond	
	n	%	n	%	n	%
Problem 1	40	30.8	87	66.9	3	2.3
Problem 2	48	36.9	80	61.5	2	1.5
Problem 3	32	24.6	95	73.1	3	2.3
Problem 4	36	27.7	90	69.2	4	3.1
Problem 5	28	21.5	98	75.4	4	3.1

The majority of all participants responded that they had not seen each problem on the MCRT before. The rate of those who responded in this manner was relatively similar across each of the problems, which is inconsistent with the rates of prior administration of these problems. The problems of the CRT have been extensively administered in prior research; however, the decoy problems of the MCRT have not. The decoy problems of the MCRT were originally given by Thomson & Oppenheimer (2016) and have not been extensively administered in previous literature. This suggests that participants were overclaiming prior exposure to the CRT problems and decoy problems.

One of the aims of including the follow up questions regarding prior exposure was to determine if prior exposure was related to performance on the MCRT problems. The following table (Table 4.1.6) provides a breakdown of the correct and incorrect responses provided by participants who reported seeing each MCRT problem before. The rows (indicated by the first column) contain each of the MCRT problems. The first column provides the total number (n) of individuals who reported seeing the MCRT problem before (this information comes directly from the first column of the preceding table, Table 4.1.5). The second column provides the number and percentage of participants who reported seeing the problem before and answered it correctly. The third column provides the number and percentage of participants who reported seeing the problem before and answered it incorrectly.

Table 4.1.6

Entire Sample MCRT Performance in Relation to Prior Exposure

MCRT Problem	Total Seen Problem Before	Correct (Out of Seen Problem Before)		Incorrect (Out of Seen Problem Before)	
	n	n	%	n	%
Problem 1	40	37	92.5	3	7.5
Problem 2	48	11	22.92	37	77.08
Problem 3	32	6	18.75	26	81.25
Problem 4	36	32	88.89	4	11.11
Problem 5	28	11	39.29	17	60.71

For each of the decoy problems (problems 1 and 4), the majority of the participants who reported seeing the problem before answered correctly. However, it should be noted that the percentage of correct responses overall to the decoy problems was observed to be high (92.3% and 86.2%, respectively). To determine if an association exists between participants' reporting that they have seen the respective decoy problems and their performance (answering correctly) on the problems, Fisher's Exact Test was performed since an assumption of the Chi-Square Test for Independence in each case was violated (one of the expected counts was less than 5). A significant association was not found between participants' reporting that they have seen problem 1 before and their performance on that problem (two-tailed significance: $p = 1.00$). A significant association was also not found for problem 4 (two-tailed significance: $p = .776$).

For each of the CRT problems (problems 2, 3, and 5), the majority of participants who reported seeing the problems before answered incorrectly. This is similar to the rate of incorrect responses given to these problems in general for all participants (13.1%, 18.5%, and 27.7%, respectively). A Chi-Square Test for Independence was conducted to determine if there exists an association between participants' reporting of prior exposure to each respective problem and their performance on the problem. Although a significant association was found for problem 2 [$\chi^2(1{,}127) = 6.191, p = .013$], the association was negative (using mean square contingency, $\varphi = -.22, V = .22$), which would imply that participants who were previously exposed to the problem were more likely to answer incorrectly to the problem. This result is likely due to the fact that the majority of all participants (86.9%) responded to the problem incorrectly. Significant associations were not found between reported prior exposure and problem performance for problem 3 [$\chi^2(1{,}127) = .001, p = .980$] and problem 5 [$\chi^2(1{,}127) = 2.765, p = .096$].

To investigate whether the association between participants' reports of prior exposure were related to one another across all problems, several Chi-Square Tests were run to produce mean square contingency coefficients (φ). The coefficients given here are equivalent to those produced through Spearman and Kendall's Tau analyses. The following table (Table 4.1.7) provides the bivariate association coefficients (φ) for each pair of responses of prior exposure to the MCRT problems. These are interpreted similarly to correlation coefficients with respect to magnitude and direction (for example, $\varphi = .814$ indicates a strong, positive association). The first column indicates each set of responses indicating prior exposure and numbers these according (the responses indicating prior exposure to Problem 1 are numbered '1.'). Then, the first column provides the number (n) of individuals who reported prior exposure to each respective problem, and the second column provides the mean response for these (in each variable, the value of 1 indicated no prior exposure, and the value of 2 indicated prior exposure).

Finally, each of the following columns are labeled according to the labels given in column 1. Constructing the table in this fashion presents each combination of bivariate associations to be displayed. The associations between a variable and itself have been omitted and replaced with hyphens. For example, in row 2, the value of .814 indicates the bivariate correlation between the responses indicating prior exposure to problem 2 (the row) and the responses indicting prior exposure to problem 1 (the column label). The asterisks given in the table indicate statistically significant associations. A single asterisk (*) indicates statistical significance at the $\alpha = .05$ level, and a double (**) asterisk indicates statistical significance at the $\alpha = .01$ level.

Table 4.1.7

Entire Sample CRT Incorrect Answer Classification

Prior Exposure Reponses	n	M	1.	2.	3.	4.	5.
1. Problem 1 Exposure	127	1.69	-	-	-	-	-
2. Problem 2 Exposure	128	1.63	.814**	-	-	-	-
3. Problem 3 Exposure	127	1.75	.700**	.720**	-	-	-
4. Problem 4 Exposure	126	1.71	.814**	.747**	.842**	-	-
5. Problem 5 Exposure	126	1.78	.579**	.614**	.653**	.634**	-

Note. A single asterisk (*) indicates significance at the $p < .05$ level. A double asterisk (**) indicates significance at the $p < .01$ level

The results of this analysis reveal that participants' reported prior exposure for each problem is strongly and positively associated with one another. This suggests that participants' who reported prior exposure to one problem tended to report prior exposure to other problems. This, coupled with the findings from the previous analysis that indicated that prior exposure is generally unrelated to increased performance, indicates that responses indicating prior exposure do not provide a reliable measure of the potential influence prior exposure may have on MCRT performance.

4.1.7 Analysis: Mathematics Course Level Subgroup MCRT Analysis

The following analyses were conducted to determine subgroup differences with respect to scores on the MCRT for students in different mathematics course levels. Recall that each mathematics course at the university under study requires the satisfaction of a prerequisite requirement. These generally take the form of certain score attained on a placement exam; the transfer of acceptable credits from another postsecondary or secondary institution; or the successful completion (often subject to minimum grade requirements) of a prerequisite course. Therefore, a direct measure of mathematical expertise is required to advance from one course level to another. Indirectly, this provides a leveling structure where courses in successive levels are separated by measurable milestones of mathematical proficiency and expertise. The university defines the academic levels in mathematics in this way according to their academic program structure.

The subgroups were formed after consulting university faculty and the university's course catalog to determine each course's prerequisite requirements; rigor and depth of the mathematical content; and whether these satisfy different subsequent quantitative major requirements. The subgroups of participants were students from introductory (commonly referred to as "remedial" or "developmental") courses in mathematics which do not satisfy the university requirements in mathematics and serve as prerequisite courses for courses in the next level (e.g., introductory/intermediate algebra); a mathematics course that satisfies the general education requirement for students not in the STEM (science, technology, engineering, or mathematics) majors (e.g., elementary statistics, mathematics for elementary education, etc.); a required mathematics course for STEM majors (e.g., calculus I, calculus II, intermediate statistics, etc.); and a major course in mathematics (e.g., foundations of mathematics, discrete mathematics, real analysis, etc.). Respectively, according to the order just described, the mathematics course level subgroups will henceforth be referred to as the "developmental level", the "general level", the "STEM level", and the "mathematics level". The number of students in each of these mathematics course level subgroups (n) is given below in Table 4.1.8 alongside their respective relative frequencies out of the total ($N = 130$) number of participants in the study.

Table 4.1.8

Descriptive Statistics of Subgroups

Subgroup	n	Relative Frequency Percentage (Out of Total 130 Participants)
All Participants	130	100%
Mathematics Level 1 (Developmental)	11	8.5%
Mathematics Level 2 (General)	55	42.3%
Mathematics Level 3 (STEM)	44	33.8%
Mathematics Level 4 (Mathematics Major)	20	15.4%

4.1.8 Analysis: Descriptive Statistics of MCRT Scores by Mathematics Course Level

Recall that one of the aims of the current study was to investigate MCRT performance for students in different levels of mathematics courses. The graphs below (Figures 4.1.9 and 4.1.10) depict the performance of participants in each mathematics course level by the number of correct answers given on the MCRT (Figure 4.1.9) and their performance by individual MCRT problem (Figure 4.1.10). The side-by-side bar graphs provide the performance for participants in each mathematics course level subgroup according to the number of correct answers given on the MCRT. Each of the bars for each respective mathematics course level subgroup depict the performance (0 correct, 1 correct, etc.) of participants in each group. The vertical axis provides a measure of the percentage of each performance level within each mathematics course level

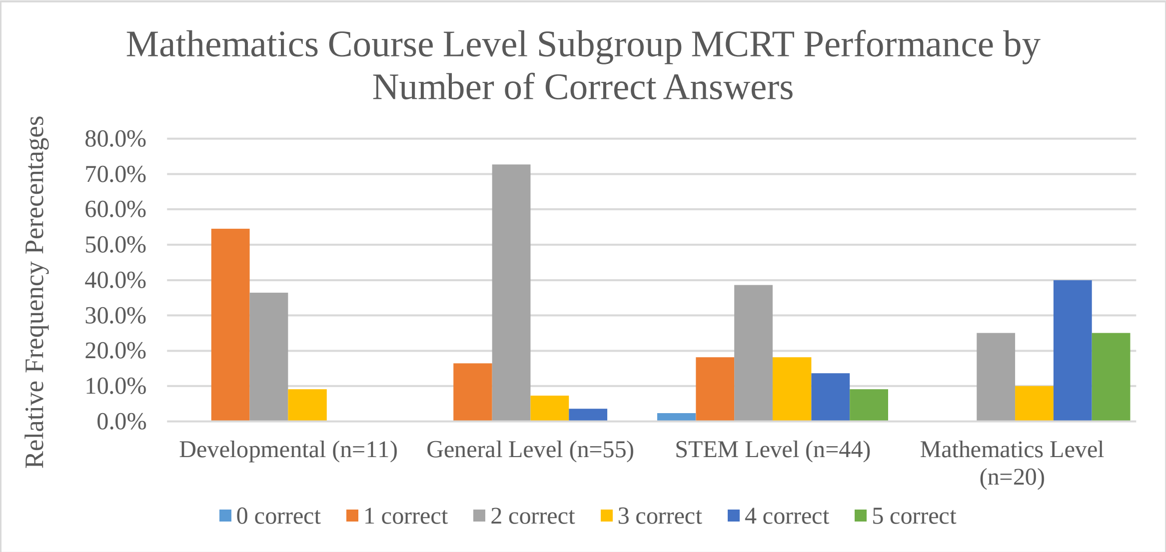

Figure 4.1.9: Graph (Side-By-Side Bar Graph) Depicting Mathematics Course Level Subgroup MCRT Performance by Number of Correct Answers

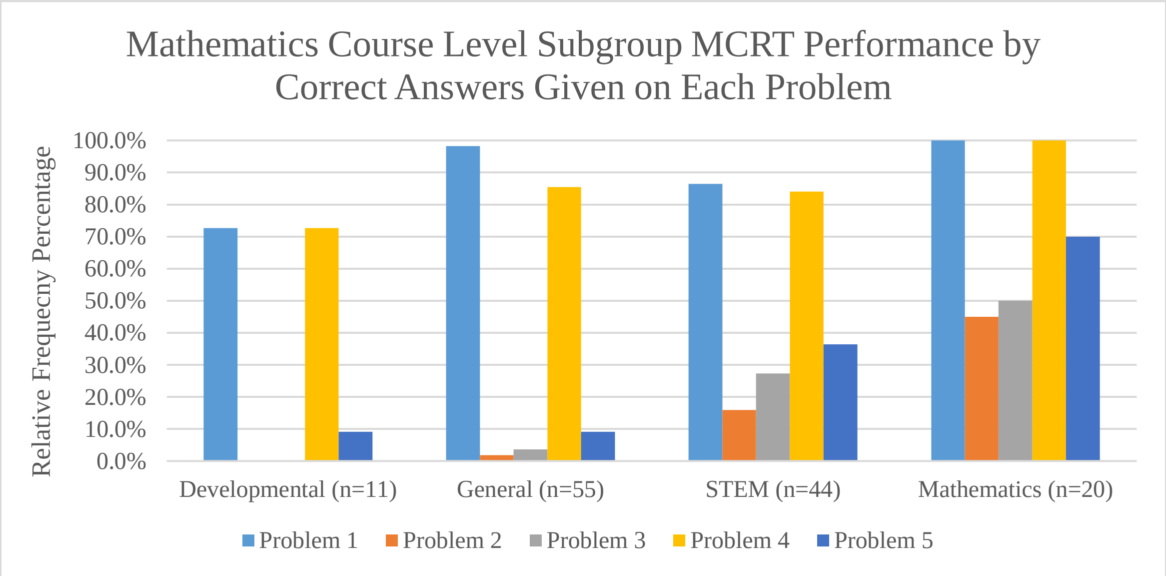

Figure 4.1.10: Graph (Side-By-Side Bar Graph) Depicting Mathematics Course Level Subgroup MCRT Performance by Percentage of Correct Answers Given on Each Problem

subgroup (out of the total number of participants in each group). The data tables used to make these graphs are given in Appendix C (Tables 4.1.9 and 4.1.10).

The most common score on the MCRT for participants in the developmental, general, and STEM level courses was 2. This is largely due to the fact that the majority of participants from each course level answered the two decoy problems correctly. Therefore, the performance on the MCRT is largely influenced by the high success rate by all participants with the decoy problems. The lowest success rate for the first and second decoy problem across all groups was 72.7% by the developmental group, and the highest was 100% by the mathematics level group.

Scores of 3 or higher on the MCRT indicate that at least one of the CRT problems were answered correctly. Participants' performance in this regard aligned with the levelling of the mathematics courses themselves. Only 1 (9.1%) student from the developmental level, 6 (10.9%) from the general level, 18 (40.9%) from the STEM level, and 15 (75%) from the mathematics level scored at least a 3 on the MCRT. This pattern also emerged at higher levels of performance on the MCRT, with none of the developmental participants earning a score of 4 or 5, only 2 (3.6%) of the general level, 10 (22.7%) from the STEM level, and 13 (65%) of the mathematics level. This pattern continued in general on each CRT problem of the MCRT, where the strongest performance was exhibited by the mathematics level students, followed by the STEM level, next by the general level, and lastly by the developmental level.

With the exception of the developmental group, across all course level subgroups, the difficulty of the CRT problems appears to match that of the entire sample. In the developmental group, no participants answered the "ball and bat" problem or the "widgets" problem correctly. For the other subgroups, the "ball and bat" problem seemed to be the most difficult, followed by the "widgets problem, followed by the "lily pads" problem. However, the respective performance on each of these problems by participants in each subgroup increased with the mathematics level of the courses in which they were enrolled. For the “ball and bat” problem, none of the developmental participants answered correctly, only 1 (1.8%) of the general students answered correctly, 7 (15.9%) of the STEM students, and 9 (45%) of the mathematics level students answered correctly. The same pattern emerged with the “widgets” problem, where it was answered correctly by none of the developmental students, 2 (3.6%) from the general level, 12 (37.3%) of the STEM level, and 10 (50%) from the mathematics level. The pattern was again observed with the final CRT problem, the “lily pads” problem, with the only exception being a tie (with respect to relative frequency percentage) between the developmental and general level groups. The “lily pads” problem was answered correctly by 1 (9.1%) of the developmental students, 5 (9.1%) of the general, 16 (36.4%) of the STEM, and 14 (70%) of the mathematics level students. In general, the strongest performance on the respective problems of the MCRT was exhibited by the mathematics level group, followed by the STEM level, then by the general level, and last by the developmental.

This pattern was consistent when viewing performance on the CRT subset of the MCRT in isolation. The following graph (Figure 4.1.11) depicts the scores on the CRT problems of the MCRT for each mathematics course level subgroup. The side-by-side bar graph provides the performance of participants in each mathematics course level subgroup according to the number of correct answers given to the CRT subset of the MCRT. Each of the bars for each respective mathematics course level subgroup depict the performance (0 correct, 1 correct, etc.) of participants in each group. The vertical axis provides a measure of the percentage of each performance level within each mathematics course level subgroup (out of the total number of participants in each group). The data table used to make these graphs are given in Appendix C (Table 4.1.11).

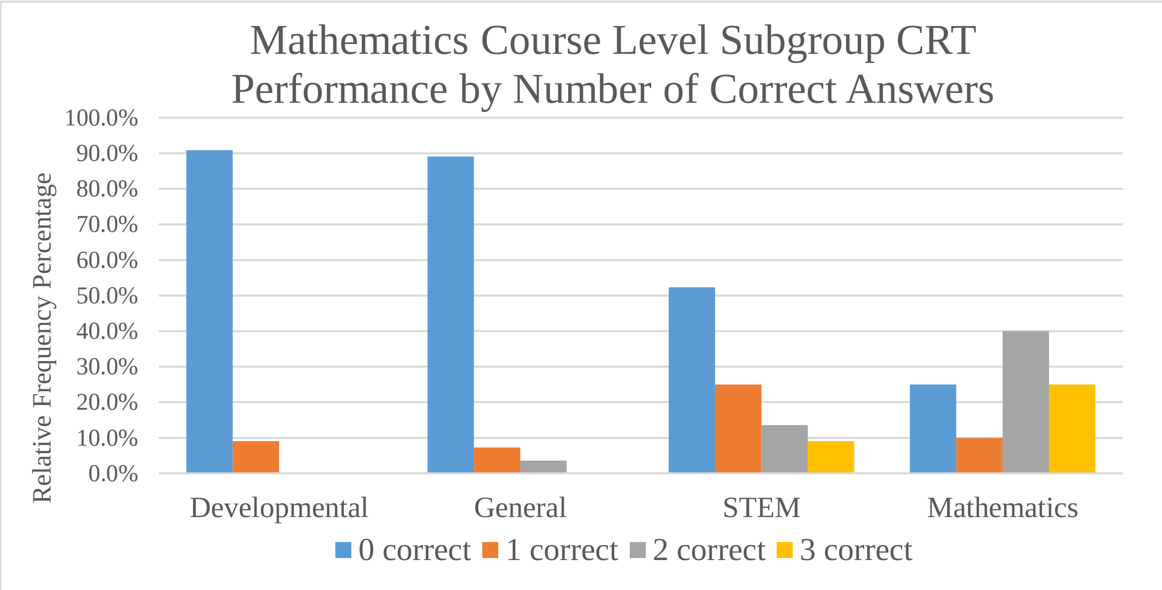

Figure 4.1.11: Graph (Side-By-Side Bar Graph) Depicting Mathematics Course Level Subgroup CRT Performance by Number of Correct Answers

The performance on the CRT in general aligns with the pattern that emerged with respect to each individual problem on the MCRT—the mathematics level students exhibited the strongest performance, followed in order by the STEM, general, and developmental level. Also, the majority of all participants from the developmental (90.9%), general (89.1%), and STEM (52.3%) course levels, respectively, earned a 0 on the CRT. Furthermore, the mathematics level was the only subgroup where the majority of the participants earned a non-zero score. If performance on the CRT was separated into two categories, zero score and non-zero score, the same pattern of performance across the subgroups emerges. The mathematics level had the highest proportion of students earning a non-zero score with 75%, followed by the STEM level with 47.7%, then by the general level with 10.9%, and lastly by the developmental level with 9.1%. This pattern was consistent when viewing the performance by subgroup with respect to a score of 1, 2, and of 3. In general, the pattern of performance that aligns with the leveling of the mathematics courses was consistent across all problems on the CRT, which was also reflected in the general performance.

However, this pattern did not emerge with respect to the analysis of the types of incorrect answers given to the CRT problems. Recall that in contrast to the decoy problems, there are two classifications of incorrect answers for the CRT problems. The two types of incorrect responses are intuitive-incorrect responses and non-intuitive-incorrect responses. For the former, these responses indicate that the solver failed to activate inhibitory control capacities, and chose the intuitive response prompted by the language and structure of the problem. For example, the intuitive-incorrect response to the "ball and bat" problem is '$.10', for the "widgets" problem it is '100 minutes', and for the "lily pads" problem it is '24 days'. The graph (Figure 4.1.12) below depicts the breakdown of incorrect responses for each of the CRT problems across each of the mathematics course level subgroups. Each bar indicates the percentage of individuals in each respective mathematics course level subgroup that provided an intuitive-incorrect response to each of the CRT problems. The vertical axis provides a measure of the relative frequency percentage of incorrect answers given by participants in each mathematics course level group (calculated out of the number of participants in each subgroup). The data table used to make this graph is given in Appendix C (Table 4.1.12).

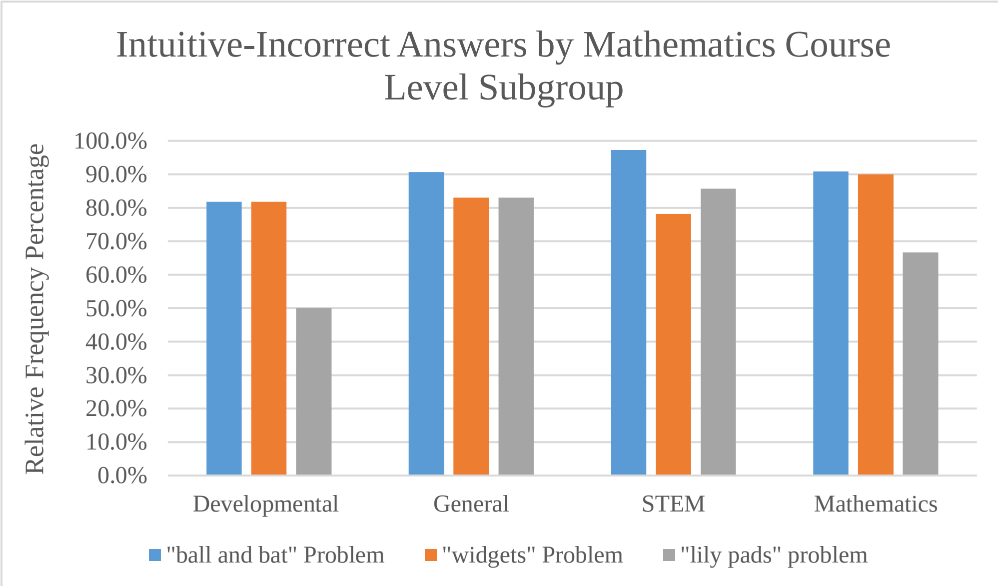

Figure 4.1.12: Graph (Side-By-Side Bar Graph) Depicting Mathematics Course Level Subgroup CRT Incorrect Answer Classification

On the “ball and bat” problem of the CRT, the vast majority of all participants in each subgroup who answered incorrectly provided the intuitive-incorrect response. In particular, 81.8% of the developmental level, 90.7% of the general level, 97.3% of the STEM level, and 90.9% of the mathematics level. It is also the case that this problem yielded the most incorrect answers on the entire MCRT. For the “widgets” problem, the vast majority who answered incorrectly again provided the intuitive-incorrect response, with 81.8% of the developmental group, 83% of the general, 78.1% of the STEM level, and 90% of the mathematics level. Participants in the mathematics level provided the greatest proportion of intuitive-incorrect answers on the widgets problem. For the final CRT problem, the “lily pads” problem, only 50% of the developmental group who answered incorrectly provided the intuitive-incorrect response, followed by 83% in the general level, 85.7% in the STEM level, and 66.7% in the mathematics level. In general, it appears that the nature of the CRT problem as default-interventionist inhibitory control tasks influenced participants' responses. Additionally, the difficulty associated with inhibiting the intuitive-incorrect response (indicated by the frequency of these responses) was experienced by participants at each mathematics course level.

4.1.9 Analysis: Inferential Statistics of MCRT Scores by Mathematics Course Level

The descriptive analysis provided in the previous section indicates that participants' performance on the MCRT and CRT subset differed with respect to their membership to different mathematics course level subgroups. In particular, the descriptive analysis suggests that not only do the subgroups differ with respect to their performance, but the differences also align with the mathematics course leveling structure. That is, the highest level of the structure performed the best, and then in descending order, the next highest level performed second best, and so on. To determine if this pattern was statistically significant, several inferential techniques were leveraged.

First, to determine if students in the mathematics course level subgroups differed from one another with respect to their scores on the MCRT and CRT, a Kruskal-Wallis Test was conducted. The Kruskal-Wallis Test was conducted instead of an Analysis of Variance (ANOVA) since the distribution of data for these variables were not normal and the sample sizes for several groups were small. Also, the Kruskal-Wallis test accounts for ties that arise in the ranked data of the both MCRT and CRT, which is reflected in the test's test statistic. A statistically significant difference was found between the scores for participants in different subgroups on the MCRT [$H(3) = 33.425, p < .001$], with a mean rank MCRT score of 101.08 for the mathematics level subgroup, 69.36 for the STEM level, 55.20 for the general level, and 36.91 for the developmental level. A statistically significant difference was also found between the scores for participants in different subgroups on the CRT [$H(3) = 39.996, p < .001$], with a mean rank CRT score of 97.55 for the mathematics level subgroup, 74.06 for the STEM level, 50.35 for the general level, and 48.77 for the developmental level.

To determine exactly which groups were significantly different from one another with respect to their scores on the MCRT and CRT, pairwise post hoc analyses utilizing Dunn's test (see IBM, 2013, p.673) were run. The Bonferonni correction was also applied to the significance values obtained from the post hoc analyses. The following tables (Tables 4.1.13 and 4.1.14) include the results from the pairwise post hoc analyses. The first group in each pairwise analysis is indicated by the label (A), and the second group is indicated by the label (B). These labels are used to denote the difference in mean rank given in the third column indicated by (A−B). If the mean rank in this third column is negative, then the mean rank for the first group exceeded that of the second group. If the mean rank of the second group exceeded that of the first group, the value in the third column indicated by (A−B) would be positive. The fourth column provides the standardized z-test statistic (which also provides the same information about the difference in the groups, where positive z-values indicate the first group's mean rank was larger, and negative z-values indicate the second group's mean rank was larger) and the fifth column provides the corresponding significance value, which has been adjusted using the Bonferonni correction for multiple tests (see the Methodology section or Dunn, 1961, 1964, IBM, 2013, p. 673, or Dinno, 2015 for more on this).

The results of the pairwise analysis for the MCRT indicate that the developmental level scored significantly lower than the STEM and mathematics level groups. The results also revealed that the students in the general level scored significantly lower than the mathematics level students.

Table 4.1.13

Post Hoc Analysis: Pairwise Comparisons of MCRT Performance by Mathematics Course Level

Pairwise Comparison				
Mathematics Course Level Subgroup (A)	Mathematics Course Level Subgroup (B)	Mean Rank Difference (A – B)	Std. Test Stat.	Sig. (Adj.)
Developmental ($n = 11$)	General ($n = 55$)	-18.29	-1.59	$.677$
Developmental ($n = 11$)	STEM ($n = 44$)	-32.44	-2.76	$.035$*
Developmental ($n = 11$)	Mathematics ($n = 20$)	-64.17	-4.89	$< .001$**
General ($n = 55$)	STEM ($n = 44$)	-14.15	-2.00	$.271$
General ($n = 55$)	Mathematics ($n = 20$)	-45.88	-5.03	$< .001$**
STEM ($n = 44$)	Mathematics ($n = 20$)	-31.72	-3.37	.005**

Note. A single asterisk (*) indicates significance at the $p < .05$ level. A double asterisk (**) indicates significance at the $p < .01$ level

Table 4.1.14

Post Hoc Analysis: Pairwise Comparisons of CRT Performance by Mathematics Course Level

Pairwise Comparison				
Mathematics Course Level Subgroup (A)	Mathematics Course Level Subgroup (B)	Mean Rank Difference (A – B)	Std. Test Stat.	Sig. (Adj.)
Developmental ($n = 11$)	General ($n = 55$)	-1.57	$-.152$	1.00
Developmental ($n = 11$)	STEM ($n = 44$)	-25.28	-2.39	$.102$
Developmental ($n = 11$)	Mathematics ($n = 20$)	-48.78	-4.14	$< .001$**
General ($n = 55$)	STEM ($n = 44$)	-23.71	-3.73	$.001$**
General ($n = 55$)	Mathematics ($n = 20$)	-47.21	-5.75	$< .001$**
STEM ($n = 44$)	Mathematics ($n = 20$)	-23.49	-2.77	.033*

Note. A single asterisk (*) indicates significance at the $p < .05$ level. A double asterisk (**) indicates significance at the $p < .01$ level

Since the pairwise analysis did not reveal a significant difference between the scores of the developmental level students and the general level students when accounting for multiple comparisons, a separate Mann-Whitney U-test revealed that the developmental level students scored significantly lower than the students in the general level [$U = 194, z = -2.25, p = .025$]. Another separate Mann-Whitney U-test was conducted to further investigate the difference between the scores of the general level students and the STEM level students, which suggests that the general level students scored significantly lower than the students in the STEM level [$U = 939, z = -2.13, p = .033$]. When taken together, the original pairwise analyses indicate that the mathematics level students exhibited the strongest performance on the entire MCRT (which includes the decoy questions and the CRT questions), and their performance was significantly stronger than all other course level groups. The combination of the separate analyses with the original pairwise analysis suggests that students in each successive level of the mathematics course leveling structure performed significantly better than those in previous levels.

The second analysis investigated participants' performance on the CRT itself, separate form the two decoy questions included on the MCRT. The results of the subgroup pairwise analysis of CRT scores revealed again that the students in the mathematics level scored significantly higher than all other course level subgroups. Also, the STEM level students scored significantly higher than the general level students. However, significant differences were not found between the performance of the developmental level students and the general and STEM level students when accounting for multiple comparisons. A separate Mann-Whitney U-test comparing the scores of the developmental students and those in the STEM level revealed that the STEM level students scored significantly better than those in the developmental level [$U = 143.5, z = -2.36, p = .018$].

When combining the subgroup analyses of scores on both the MCRT and CRT, the results suggest that students in higher mathematics course levels performed better than those at lower levels. In particular, students in the mathematics level scored significantly higher than students in every other level on both measures. The results also suggest that the performance on both measures increased according to the mathematics course leveling structure. That is, the participants in the mathematics level had the highest scores, the participants in the STEM level had the second highest scores, the participants in the general level had the third highest scores, and the participants in the developmental level had the lowest scores. In essence, students in courses that ranked higher on the course leveling structure performed better, in general, than students in lower levels, and that many of these stronger performances were statistically significant.

4.1.10 Analysis: Additional Subgroup Analyses

Although the main subgroups under analysis were those formed by the type of mathematics courses in which participants were enrolled, other subgroups within the sample were identified and further investigated. Recall that participants were intentionally recruited from different levels of mathematics courses at the university under study by design. Therefore, membership to the mathematics course level subgroups was determined at the onset of the study. Membership to additional subgroups in the current analysis was determined by participants' responses to the demographic and grouping questions included at the beginning of the SMPI (psychoeducational survey). The subgroups that were formed in this manner according to participants' self-reported gender (male and female, since category of "other" or "prefer not to say" only included 2 participants), academic year (first-year, sophomore, junior, and senior), and academic major (separated into majors in health and human services; education; sciences; business; and undeclared or no major). Although a number of demographic and subgroup questions were included at the beginning of the psychoeducational survey, only those related to the subgroupings mentioned above yielded actionable information in analysis. Other demographic questions on the SMPI did not yield data that were suitable for further analysis, and therefore, these were omitted from subgroup analyses (for a discussion of this, see Chapter 5).

A breakdown of the number of participants within each subgroup undermeasure is given below. The number of participants in each subgroup (n) is given alongside the relative frequency of subgroup membership with respect to the entire sample (out of 130 participants) is given. Note that the groups given below are not disjoint; that is, a participant may belong to one or more groups. For example, a participant may be a female, a sophomore, and in the education-related majors. The following table (Table 4.1.15) below provides the number of participants belonging to each subgroup and the relative frequency percentage of each group out of the total number of participants (130), respectively.

4.1.11 Analysis: MCRT Scores by Gender

For the question that asked participants to report their gender, 78 reported being female, 49 reported being male, 1 participant selected the option of "Prefer not to say", 1 participant reported that they identified as "non-binary", and 1participant did not provide a response. There were two main subgroups that emerged, male participants and female participants. Since there was only one participant that reported a gender other than male or female, subgroup analysis with respect to gender including this individual in a separate group was not conducive to statistical analysis. Therefore, the three students that did not report being male or female were excluded from the analysis, and the subgroup analysis with respect to gender was focused on the two large subgroups of participants who reported being male and female. Such analyses have been conducted many times in previous research on CRT performance (Avram, 2018; Campitelli & Gerrans, 2014; Frederick, 2005; Brañas-Garza et al., 2019).

Table 4.1.15

Descriptive Statistics of Subgroups

Subgroup	n	Relative Frequency Percentage (Out of Total 130 Participants)
All Participants	130	100%
Gender: Male	49	37.7%
Gender: Female	78	60%
Gender: Other	2	1.6%
Academic Year: First-Year	46	35.7%
Academic Year: Sophomore	26	20.2%
Academic Year: Junior	34	26.4%
Academic Year: Senior	22	17.1%
Academic Year: Other (Not Specified)	1	.8%
Academic Majors: Business-Related Majors	5	3.8%
Academic Majors: Education-Related Majors	15	11.5%
Academic Majors: Health and Human Services-Related Majors	35	26.9%
Academic Majors: Mathematics Majors	22	16.9%
Academic Majors: Science-Related Majors	47	36.2%
Academic Majors: No Major Declared	6	4.6%
High School Mathematics Course Grouping: None	11	8.5%
High School Mathematics Course Grouping: Algebra	20	15.4%
High School Mathematics Course Grouping: General	25	19.2%
High School Mathematics Course Grouping: Precalculus	32	24.6%
High School Mathematics Course Grouping: Advanced	30	23.1%

Note. Individual participants may belong to multiple subgroups (for example, a participant may be in the female, sophomore, mathematics major, and advanced high school course subgroups).

With respect to their scores on the MCRT, the majority of both males (53.1%) and females (80.8%) scored a 2 on the MCRT. This is heavily influenced by the high success rate exhibited by both groups on the first (89.8% of males correctly answered, 93.6% of females correctly answered) and second (91.8% of males correctly answered, 83.3% of females correctly answered) decoy questions of the MCRT.

The graph below (Figure 4.1.15) depicts the performance of each gender subgroup on the CRT subset of the MCRT. This side-by-side bar graph provides the performance of each subgroup on the CRT according to the number of correct answers given. Each bar indicates the relative frequency percentage of each type of performance in each of the subgroups (calculated out of the total number of participants in each subgroup).

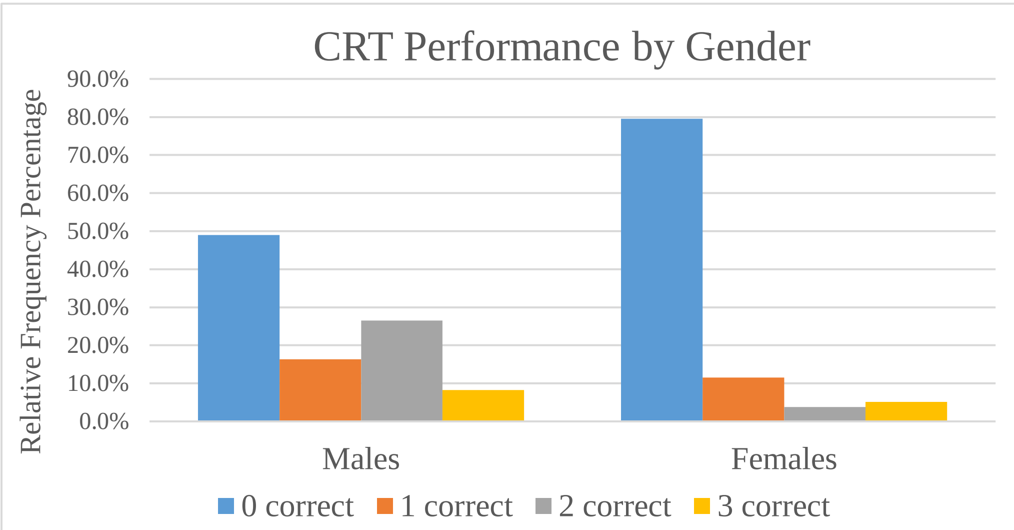

Figure 4.1.16: Graph (Side-By-Side Bar Graph) Depicting CRT Performance by Gender

A descriptive analysis of the CRT questions of the MCRT revealed that males scored better on the “ball and bat” problem (18.4% of males answered correctly compared to 9% of females), the “widgets” problem (32.7% of males answered correctly compared to 7.7% of females), and on the “lily pads” problem (42.9% of males answered correctly compared to 17.9% of females). The majority of the males (51%) earned a non-zero score on the CRT, whereas the vast majority of females (79.5%) earned a zero score. In terms of the types of incorrect answers given on to problems of the CRT, the vast majority of both groups provided the intuitive-incorrect answers on each problem: for the “ball and bat” problem, 95% of the incorrect answers provided by males were intuitive incorrect answers compared to 88.9% of females’; for the “widgets “ problem, 91% of incorrect answers provided by males were intuitive-incorrect answers compared to 77.8% of females’; and for the “lily pads” problem, 75% of incorrect answers provided by males were intuitive-incorrect answers compared to 85.9% of females'.

To determine if males and females differed significantly from one another with respect to their scores on the MCRT and CRT, a set of Mann-Whitney U-tests were conducted. A statistically significant difference was found between males and females scores on the MCRT [$U = 1363.5, z = -2.94, p = .003$], with males (mean rank: 75.17) scoring significantly higher than females (mean rank: 56.98). A statistically significant difference was also found between males and females scores on the CRT [$U = 1299.5, z = -3.66, p < .001$], with males (mean rank: 76.48) scoring significantly higher than females (mean rank: 56.16). Taken together, the results of these analyses suggest that males performed better than females on both the MCRT and CRT, with stronger performances on each individual problem of the CRT and in terms having a non-zero score on the CRT.

4.1.12 Analysis: MCRT Scores by Academic Year

For the question that asked participants to report their academic year, 46 reported being in their first-year, 26 reported being in their sophomore year, 34 reported being in their junior year, 22 reported being in their senior year, 1 reported not being in any academic year, and 1 did not respond to this question. Since there was only one participant that reported not being in any academic year, subgroup analysis with respect to academic year including this individual in a separate group was not conducive to statistical analysis. Therefore, the student who did not respond to this question and the student who reported not being in any academic year were excluded from the analysis.

With respect to their performance on the MCRT, the majority of first-year students (65.2%) and sophomores (53.8%) scored a 2 on the MCRT. This was followed by 44.1% of juniors and 31.8% of seniors scoring a 2 on the test. As with previous analyses, this is heavily influenced by the fact that the vast majority of all participants in each individual academic year subgroup correctly answered the first decoy problem (97.8% of first-year students, 84.6% of sophomores, 91.2% of juniors, and 90.9% of seniors) and the second decoy problem (87% of first-year students, 80.7% of sophomores, 91.2% of juniors, and 86.4% of seniors). A descriptive analysis of CRT performance indicated that at least half of all participants in each group earned a zero score (80.4% of first-year students, 73.1% of sophomores, 55.9% of juniors, and 50% of seniors). Moreover, the percentage of individuals not earning any points on the CRT decreased steadily with each successive academic year.

The graph below (Figure 4.1.16) depicts the performance of each academic year subgroup on the CRT subset of the MCRT. This side-by-side bar graph provides the performance of each subgroup on the CRT according to the number of correct answers given. Each bar indicates the relative frequency percentage of each type of performance in each of the subgroups (calculated out of the total number of participants in each subgroup).

A descriptive analysis of the CRT questions of the MCRT revealed that juniors performed best on the "ball and bat" problem (23.5% of juniors answered correctly, followed by 22.7% of seniors, 7.7% of sophomores, and 2.2% of first-year students) and on the "widgets" problem (29.4% of juniors answered correctly, followed by 22.7% of seniors, 13% of first-year students, and 7.7% of sophomores). For the "lily pads" problem, seniors had the strongest performance (50% of seniors answered correctly, followed by 29.4% of juniors, 26.9% of sophomores, and 15.2% of first-year students). In terms of the types of incorrect answers given on each of these CRT problems, the vast majority of individuals in each group provided the incorrect-intuitive response on each CRT problem.

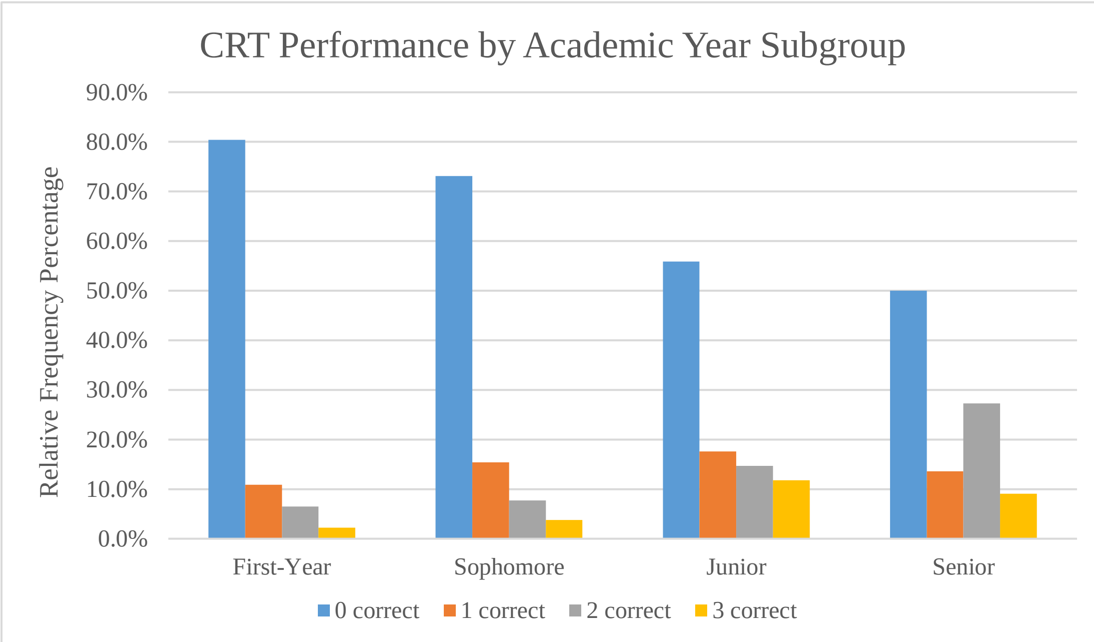

Figure 4.1.17: Graph (Side-By-Side Bar Graph) Depicting CRT Performance by Academic Year Subgroup

For the "ball and bat" problem, 87% of first-year students provided the intuitive-incorrect response, followed by 80.8% of sophomores, 76.5% of juniors, and 72.7% of seniors. For the "widgets" problem, 84.6% of sophomores provided the incorrect-intuitive response, followed by 69.6% of first-year students, 58.8% of juniors, and 54.5% of seniors. Lastly, for the "lily pads" problem, 73.9% of first-year students provided the incorrect-intuitive response, followed by 58.8% of juniors, 57.7% of sophomores, and 36.4% of seniors.

To determine whether at least one of the academic year subgroups differed significantly from one another with respect to their scores on the MCRT and CRT, a set of Kruskal-Wallis tests were conducted. A statistically significant difference was not found between the scores of students from different academic years on the MCRT [$H(3) = 6.942, p = .074$]. However, a statistically significant difference was found between the scores of students from different academic years on the CRT subset of the MCRT [$H(3) = 10.246, p = .017$]. This analysis also revealed that the man rank CRT score for each subgroup were 55.41 for first-year students, 59.92 for sophomores, 72.35 for juniors, and 76.77 for seniors. A subsequent pairwise analysis with a Bonferonni correction revealed that first-year students scored significantly lower than seniors on the CRT [mean rank difference: -$21.36, z = -1.885, p = .045$]. Although a significant difference was not found between first-year students and juniors when accounting for all comparisons, a separate Mann-Whitney U-test was conducted and indicated that juniors scored higher than first-year students on the CRT [$U = 576, z = 2.481, p = .013$]. Taken together, the results suggest that performance on the CRT increases as according to academic year, and that this relationship was statistically significant when comparing first-year students and upper-class students.

4.1.13 Analysis: MCRT Scores by Academic Major

For the question that asked participants to report their academic major, all 130 participants responded. From these responses, six groupings were created (see Table 4.1.15 for a breakdown of the size of each group): business-related majors (e.g., business administration, business in finance, business management, etc.), education-related majors (e.g., early childhood education, elementary education, collaborative education, special education, etc.), health and human services-related majors (e.g., public health, nursing, healthcare studies, psychology, exercise science, social work, etc.), mathematics-related majors (e.g., mathematics, applied mathematics, mathematics with secondary education certification, etc.), science-related majors (e.g., physics, computer science, biochemistry, biology, etc.), and students with no majors (e.g., undeclared, undecided, etc.).

With respect to their performance on the MCRT, the majority of all participants in each of the academic major subgroups, except mathematics, scored at most a 2 on the MCRT. This was again heavily influenced by the success rate on each of the decoy problems. The mathematics major subgroup was the only subgroup with the majority of participants (68.2%) scoring above a 2. With respect to performance on the CRT, the participants in the mathematics major subgrouping had the strongest performance. Only 22.7% of the mathematics major grouping scored a 0 on the CRT, whereas the vast majority of all participants in each of the other academic major subgroupings scored a 0 (100% of the students with no majors, followed by 93.3% of education-related majors, followed by 88.6% of health and human services-related majors, 60% of business-related majors, and 59.6% of science majors). Only the major groupings of mathematics and science contained participants who earned a perfect score (3) on the CRT (27.3% of mathematics and 6.4% of science).

The graph below (Figure 4.1.17) depicts the performance of each academic major subgroup on the CRT subset of the MCRT. This side-by-side bar graph provides the performance of each subgroup on the CRT according to the number of correct answers given. Each bar indicates the relative frequency percentage of each type of performance in each of the subgroups (calculated out of the total number of participants in each subgroup).

On each CRT problem, students in the mathematics-related majors earned the highest scores (45.5% correct on the "ball and bat" problem, 50% on the "widgets" problem, and 72.7% on the "lily pads" problem), followed by students in the science-related majors (12.8% correct on the "ball and bat" problem, 23.4% on the "widgets" problem, and 27.7% on the "lily pads" problem). It should be noted that 2 out of 6 (60%) of business-related majors correctly answered the "lily pads" problem as well. On the other end of the spectrum, 100% of the students with no major answered each of the CRT questions incorrectly.

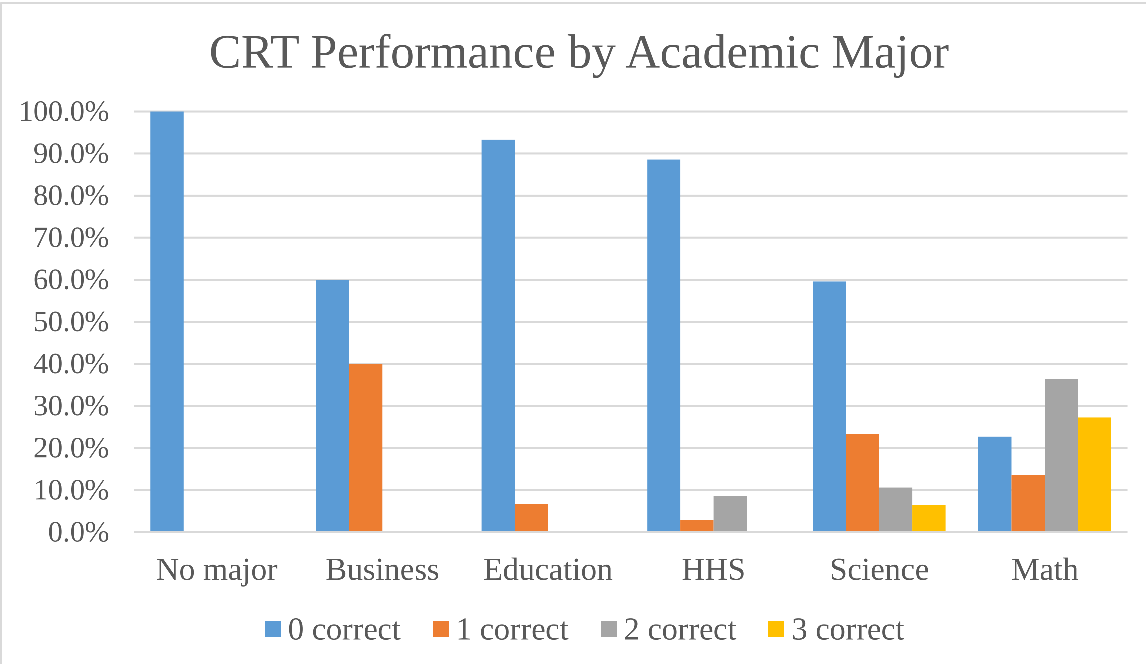

Figure 4.1.18: Graph (Side-By-Side Bar Graph) Depicting CRT Performance by Academic Major Subgroup

Also, 100% of participants in the education-related majors incorrectly answered the "ball and bat" and "widgets" problems, and 93.3% answered the "lily pads" problem incorrectly. The health and human services majors followed next, with only 2.9% correctly answering the "ball and bat" problem, 5.7% correctly answering the "widgets" problem, and 11.4% answering the "lily pads" problem correctly. In terms of the types of incorrect answers given by participants in each major subgroup, at least half of all participants in each group provided the intuitive-incorrect response to the "ball and bat problem" and the "widgets" problem. With the exception of the mathematics and business majors, this was also true for the "lily pads" problems.

To determine whether at least one of the academic year subgroups differed significantly from one another with respect to their scores on the MCRT and CRT, a set of Kruskal-Wallis tests were conducted. A statistically significant difference was found between the scores of students from different academic majors on the MCRT [$H(5) = 26.48,\ p < .001$]. The mean ranks calculated in this analysis were 96.36 for mathematics-related majors, 67.67 for science-related majors, 55.83 for health and human services-related majors, 51.30 for education related majors, 47 for business-related majors, and 42.67 for students with no major. Pairwise post hoc analyses with a Bonferonni correction revealed that students in the mathematics major scored significantly higher than students in each of the other academic majors, with the exception of business (this finding could be due to the fact that the business major subgroup contained only the minimum number of 5 participants for this analysis, since a separate Mann-Whitney U-test for this difference was statistically significant with $U = 16.5, z = -2.47, p = .014$). That is, mathematics majors had statistically significant, higher scores on the MCRT than: students with no major [mean rank difference: $53.7, z = 3.337, p = .013$], students in the education-related majors [mean rank difference: $45.06, z = 3.85, p = .002$], students in the health and human services-related majors [mean rank difference: $40.54, 4.27, p < .001$], and students in the science-related majors [mean rank difference: $28.69, z = 3.18, p = .008$].

A statistically significant difference was also found between the scores of students from different academic majors on the CRT subset of the MCRT [$H(5) = 40,\ p < .001$]. The mean ranks calculated in this analysis were 98.80 for mathematics-related majors, 68.91 for science-related majors, 65 for business-related majors, 51.46 for health and human services-related majors, 47.5 for education related majors, and 44 for students with no major. Pairwise post hoc analyses with a Bonferonni correction revealed that students in the mathematics major scored significantly higher than students in each of the other academic majors, with the exception of business (this finding could again be due to the fact that the business major subgroup contained only the minimum number of 5 participants for this analysis, since a separate Mann-Whitney U-test for this difference was statistically significant with $U = 20.5, z = -2.231, p = .026$). That is, mathematics majors had statistically significant, higher scores on the MCRT than: students with no major [mean rank difference: $54.75, z = 3.79, p = .002$], students in the education-related majors [mean rank difference: $51.3, z = 4.88, p < .001$], students in the health and human services-related majors [mean rank difference: $47.34, z = 5.54, p < .001$], and students in the science-related majors [mean rank difference: $29.881, z = 3.682, p = .003$]. Taken together, the results of these analyses indicate that students in the mathematics major performed better than other majors on the CRT in general.

4.1.14 Analysis: MCRT Scores by Mathematics Course Taken Senior Year of High School

For the question that asked participants to report whether or not they took a math class their senior year of high school, all 130 participants responded. Out of the total 130 participants, 118 indicated that they took a math class their senior year of high school and 11 did not. If respondents reported taking such a course, a subsequent question asked them to report which class hey took. Of the 118 participants that reported taking a math class their senior year of high school, 12 either did not answer the subsequent question asking them to specifying the course that they took or could not recall (e.g., responses included “I don’t remember” or “I’m not sure”). Therefore, these participants were not included in the analyses conducted on subgroups formed based upon the type of mathematics class taken in their senior year of high school.

Based upon the responses to these questions, five subgroups were formed. The first group comprised the participants that did not take a math class their senior year of high school. The second group included participants that reported taking an algebraic course their senior year. The titles of these courses differed greatly, however, if the title included the word algebra or was known to primarily cover algebra (based upon regional knowledge of high school curriculum) it was included in this group. Course titles that were selected for this group included Algebra,

Algebra II, Algebra III, Algebra and Statistics, ALEKS Program, College Algebra, Introduction to College Mathematics, Pre-Algebra, and Trigonometry and Algebra. The second group included participants that reported taking courses in precalculus. Course titles that were selected for this group included precalculus, advanced algebra and precalculus, precalculus and trigonometry, and precalculus and statistics. The third group included participants that reported taking advanced courses. Course titles that were elected or this group included Calculus, AP Calculus, AP Statistics, and Multivariable Calculus.

The final group included participants that reported taking courses with titles that were not aligned with the other groups and differed greatly from one another. This grouping also included courses are non-traditional when considering graduation requirements and trends in high schools in the northeast. These courses included accounting, general statistics, basic statistics, probability, and business mathematics. It is important to note that this final group does not have a well-defined structure. The courses that were included could be advanced in nature given their course titles and could have been grouped differently. However, since the numbers of participants taking well-known courses generally defined courses (e.g., statistics) were low and would not form groupings conducive to statistical analysis, they were combined to form this general group. It is noted here that they way in which this grouping was formed may confound any findings gleaned from analyses with this group. The size of each subgroup (no high school math course, algebra, precalculus, advanced courses, and general courses) is given in Table 4.1.15.

With respect to their performance on the MCRT, the majority of all participants in each of the high school course subgroupings, except those in the advanced course grouping, scored at most a 2 on the MCRT. This was heavily influenced by the success rate on the decoy problems.

The participants who took advanced courses in high school were the only one with the majority of participants (63.3%) scored above a 2, this was followed by 36% in the no high school math course grouping, 21.9% from the precalculus grouping, 20% from the general grouping, and 5% in the algebra grouping. With respect to performance on the CRT subset of the MCRT, the advanced course subgrouping exhibited the strongest performance, with 66.7% earning non-zero scores and 16.7% earning perfect score. The algebra subgrouping had the lowest performance, and accordingly the highest percentage of participants scoring a zero on the CRT with 95%, followed by 80% of the general subgrouping, 75% of the precalculus subgrouping, and 63.6% of the no high school math course subgrouping.

The graph below (Figure 4.1.18) depicts the performance of each high school mathematics course level subgroup on the CRT subset of the MCRT. This side-by-side bar graph provides the performance of each subgroup on the CRT according to the number of correct answers given. Each bar indicates the relative frequency percentage of each type of performance in each of the subgroups (calculated out of the total number of participants in each subgroup).

On each CRT problem, participants in the advanced course subgroupings had the highest scores (30% correct on the “ball and bat” problem, 46.7% on the “widgets” problem, and 53.3% on the “lily pads” problem). On the other end of the spectrum, 100% of the participants in the algebra subgrouping answered the “ball and bat” and “widgets” problems incorrectly, and 95% of these participants answered the “lily pads” problem incorrectly. For the remaining subgroupings, only 9.1% of the no high school math course grouping, 8% of the general grouping, and 12.5% of the precalculus grouping answered the “ball and bat” problem correctly;

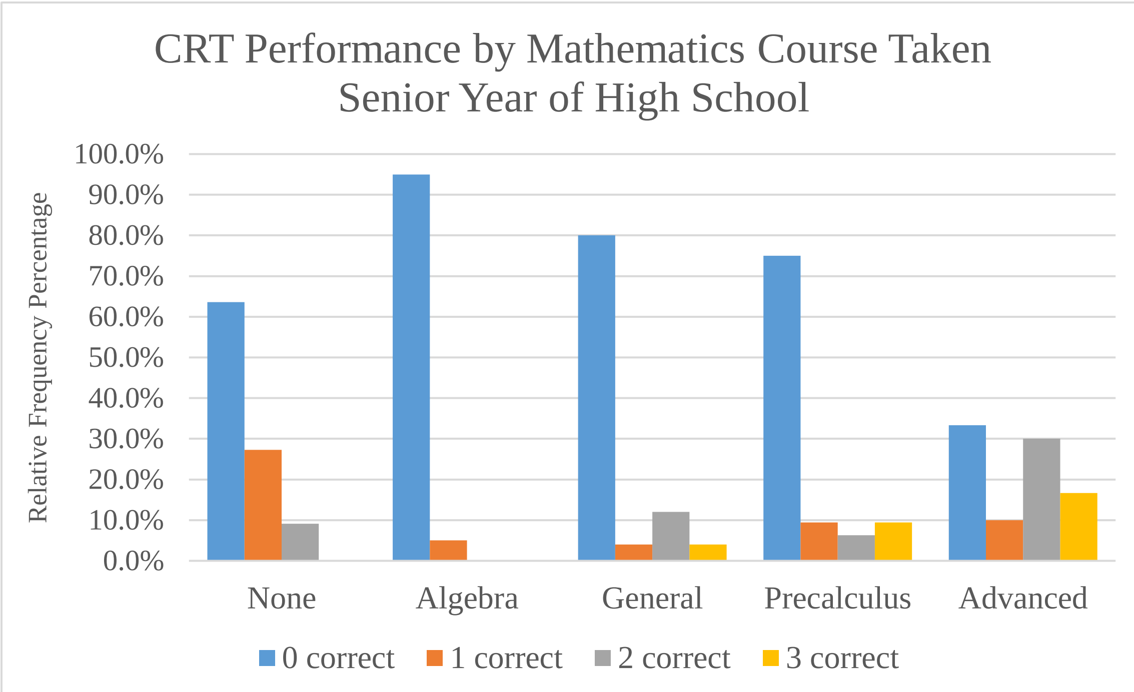

Figure 4.1.19: Graph (Side-By-Side Bar Graph) Depicting CRT Performance by Mathematics Course Taken Senior Year of High School

only 9.1% of the no high school math course grouping, 16% of the general grouping, and 15.6% of the precalculus grouping answered the "widgets" problem correctly; and 27.3% of the no high school math course grouping, 16% of the general grouping, and 21.9% of the precalculus grouping answered the "lily pads" problem correctly. In terms of the types of incorrect answers given by participants, the vast majority of all participants in each subgroup except the advanced course level group (highest percentage being 90.9% in the algebra subgroup and lowest being 62.5% in the no high school math course group) provided the intuitive-incorrect answers when considering all answers provided (both correct and incorrect). In the advanced course level subgroup, the majority of the incorrect answers given were the intuitive-incorrect answers.

To determine whether at least one of the high school math course subgroups differed significantly from one another on the MCRT and CRT, a set of Kruskal-Wallis tests were conducted. A statistically significant difference was found between the scores of the participants form different high school math course subgroupings on the MCRT [$H(4) = 24.68,\ p < .001$]. The mean ranks calculated in this analysis were 82.07 for the advanced course subgroup, 55.63 for the precalculus subgroup, 55.46 for the general, 55.45 for the no high school math course subgroup, and 39.13 for the algebra subgroup. Pairwise post hoc analyses with a Bonferonni correction revealed that the participants in the advanced level course subgroup scored significantly higher than participants in each of the other high school math course subgroups, with the exception of the group of individuals who reported not taking a math course their senior year of high school (it should be noted that a separate Mann-Whitney U-test for this difference was statistically significant with $U = 86.5, z = -2.39, p = .019$). In particular, the advanced course level subgroup scored higher than the participants in the algebra subgroup [mean rank difference: $42.942, z = 4.707, p < .001$], the general course subgroup [mean rank difference: $26.607, z = 3.109, p = .019$], and the precalculus course subgroup [mean rank difference: $26.442, z = 3.292, p = .010$].

A statistically significant difference was also found between the scores of participants from different high school math course subgroups on the CRT subset of the MCRT [$H(4) = 25.369,\ p < .001$]. The mean ranks calculated in this analysis were 80.60 for the advanced course subgroup, 55.64 for the precalculus subgroup, 52.70 for the general, 58.91 for the no high school math course subgroup, and 42.85 for the algebra subgroup. Pairwise post hoc analyses with a Bonferonni correction revealed that the participants in the advanced level course subgroup scored significantly higher than participants in each of the other high school math course subgroups, with the exception of the group of individuals who reported not taking a math course their senior year of high school (it should be noted that a separate Mann-Whitney U-test for this difference was statistically significant with $U = 94.5, z = -2.184, p = .037$). In particular, the advanced course level subgroup scored higher than the participants in the algebra subgroup [mean rank difference: $37.75, z = 4.261, p < .001$], the general course subgroup [mean rank difference: $27.9, z = 3.641, p = .003$], and the precalculus course subgroup [mean rank difference: $24.959, z = 3.471, p < .001$]. Taken together, the results of these analyses indicate that the participants who took advanced level mathematics courses their senior year of high school performed better than those who took lower level courses in general.

4.2 Research Question 2: How do students from different collegiate mathematics courses respond to a psychoeducational survey measuring domain-general academic habits of mind, future orientation, and academic self-concept, and mathematics-specific academic habits of mind, beliefs, and past academic experiences?

4.2.1 Analysis: SMPI Item Analysis

As previously mentioned all 130 participants in the current study responded to the items of the SMPI. Also, as previously mentioned, the first seven items on the SMPI were demographic in nature and aided in the identification of subgroups for analysis (see Appendix B for the statements of these questions and for the full SMPI). These included questions asking for participants to report their academic year; their academic major; whether the class they were taking was a prerequisite for another course, one that satisfied their general education requirement, or was an elective; whether the students placed into the course out of high school, satisfied a prerequisite requirement for the course, or another reason; and whether they took a mathematics course their senior year of high school (if they did take a course their senior year of high school, a follow up question asked them to report which course they took.

In viewing the responses, it seems that the question that asked students about the how the mathematics course in which they were enrolled related to their degree requirements (whether the course was a prerequisite, satisfied their general education requirement, or was an elective) was flawed. This question did not have a follow up question that provided participants with the opportunity to write in a response that differed from the options that were presented. Several students wrote in such responses near the question itself, which indicated that the question did not account for all potential responses, and as a result, the data from this question was potentially inaccurate. Therefore, it was omitted from further analysis (see Chapter 5 for a discussion of this).

Responses to the question regarding the nature of participants' enrollment in the course were also omitted from further analysis. This question asked participants whether they were placed into the course out of high school, whether they satisfied a prerequisite requirement, or another reason. This question did have an option for participants to write in their responses if they differed from the options that were presented. These write-in responses indicated that the question did not account for other circumstances related to enrollment, and therefore the data may be inaccurate. Thus, subgroup formation and responses related to this question were omitted from the analysis (see Chapter 5 for a discussion of this). For ease of viewing the descriptive statistics regarding subgroup membership for all subgroups under analysis are given in Table 4.2.1 below.

The remaining items of the SMPI included 33 Likert scale items measuring level of agreement (1=strongly disagree, 2=disagree, 3=neutral, 4=agree, 5=strongly agree). Each of these items measured indicators of psychoeducational facets related to domain-general academic habits of mind, future orientation, and academic self-concept, and mathematics-specific academic habits of mind, beliefs, and past academic experiences.

Table 4.2.1

Descriptive Statistics of Subgroups

Subgroup	N	Relative Frequency Percentage (Out of Total 130 Participants)
All Participants	130	100%
Mathematics Level 1 (Developmental)	11	8.5%
Mathematics Level 2 (General)	55	42.3%
Mathematics Level 3 (STEM)	44	33.8%
Mathematics Level 4 (Mathematics Major)	20	15.4%
Gender: Male	49	37.7%
Gender: Female	78	60%
Gender: Other	2	1.6%
Academic Year: First-Year	46	35.7%
Academic Year: Sophomore	26	20.2%
Academic Year: Junior	34	26.4%
Academic Year: Senior	22	17.1%
Academic Year: Other (Not Specified)	1	.8%
Academic Majors: Business-Related Majors	5	3.8%
Academic Majors: Education-Related Majors	15	11.5%
Academic Majors: Health and Human Services-Related Majors	35	26.9%
Academic Majors: Mathematics Majors	22	16.9%
Academic Majors: Science-Related Majors	47	36.2%
Academic Majors: No Major Declared	6	4.6%
High School Mathematics Course Grouping: None	11	8.5%
High School Mathematics Course Grouping: Algebra	20	15.4%
High School Mathematics Course Grouping: General	25	19.2%
High School Mathematics Course Grouping: Precalculus	32	24.6%
High School Mathematics Course Grouping: Advanced	30	23.1%

Note. Individual participants may belong to multiple subgroups (for example, a participant may be in the female, sophomore, mathematics major, and advanced high school course subgroups).

The final question was an open-ended question asking participants "Is there anything else you would like us to know?" A large space was provided for participants to write in their responses. In the succeeding sections, descriptive statistics for all 33 Likert scale SMPI items are given for each of the subgroups under analysis. These are all given with the percent disagreement to each item (calculated by summing the responses of disagree and strongly disagree and dividing by the total number of responses for each question) and the percent agreement to each item (calculated by summing the responses of agree and strongly agree and dividing by the total number of responses for each question).

4.2.2 Analysis: Descriptive Statistics of SMPI Items Across Entire Sample

The percentage agreement and disagreement for all 33 Likert scale items across the entire sample are included in Table 4.2.2 below. Only the percentage agreement (calculated by combining the responses of "agree" and "strongly agree" for each item) and disagreement (calculated by combining the responses of "disagree" and "strongly disagree" for each item) are displayed for ease of viewing. Grouping responses in this way also provides an overall measure of each type of sentiment to each item. The "neutral" responses were not reported; however, since only the relative frequencies of responses indicating disagreement and agreement are given, the remaining percentage (that is, subtracting the sum of percentages indicating disagreement and agreement away from 100%) indicates the percentage of responses of "neutral".

All 130 participants responded to the SMPI including the section consisting of the 33 Likert scale items. Missing data on these 33 items was exceptionally low—there were only 7 missing responses in total, on 6 of the total 33 items. The item "I tell my professor when I don't understand something from math class", only had 2 missing responses. The item "The hardest part about solving word problems is understanding what is being asked" had only 1 missing response. The item "Thinking about the future I want makes me do more now to get that future" only had 1 missing response. The item "I wait until right before a math test to start studying" only had 1 missing response.

Table 4.2.2

Percentage Agreement and Disagreement to All Items on the SMPI Across the Entire Sample

SMPI Likert Scale Item	D%	A%
I'm just not good at math.	46.9	31.5
The wording of math problems confuses me.	19.2	53.1
I tell my professor when I don't understand something from math class.	21.1	56.3
The hardest part about solving word problems is understanding what is being asked.	18.6	69
I use my time between classes productively.	15.4	60
I settle for just passing my courses.	52.3	27.7
I break down long-term assignments and/or class projects and work on them over time.	37.7	38.5
I often play catch-up in my classes.	41.5	37.7
I have a fairly clear idea of what I need to study now in order to have the career I want.	8.5	70.8
Math and/or anything with numbers has been an obstacle to my academic success.	53.1	32.3
Thinking about the future I want makes me do more now to get that future.	3.9	81.4
I expect to use the math I have learned in my future career.	26.2	55.4
I have experienced difficulties in math since high school or before high school.	40	51.5
I am always well-prepared for math class.	19.2	43.8
I am usually confident that I will do well on math tests.	35.4	34.6
I wait until right before a math test to start studying.	44.6	26.9
I have a "game plan" that is effective for tackling math homework.	29.2	47.7
I give myself enough time to really read course materials.	25.4	43.8
I push aside math assignments and do them last.	51.5	28.5
I am taking the necessary steps to obtain the career I desire.	1.5	89.2
I can explain how I reach the correct answer on a math test.	10	67.7
When something confuses me, I think about it until I can make sense out of it.	6.2	82.3
I find it hard to prioritize my time.	29.2	46.2
My confidence in academic skills has increased this semester.	22.3	45.4
I study regularly to be successful in college.	19.2	56.2
Sometimes, I am disappointed in my test results because I studied a great deal.	20.8	61.5
Other commitments in my life get in the way of studying for classes.	15.6	68.8
My grades in college math have influenced what degree I can pursue.	31.5	34.6
I am hesitant to raise my hand in class even though I know the answer.	33.1	50.8
I fear that if I ask for help, my professor will think less of me.	61.5	23.8
I am definitely a "work before play" type of person.	21.5	40
I often get distracted during class by feelings of boredom.	26.2	50.8
I am doing better than I thought I would in college.	19.2	46.2

Note. D% represents the percent of responses that indicated disagreement to each item (responses of disagree or strongly disagree), and A% represents the percent of responses that indicated disagreement to each item (responses of disagree or strongly disagree).

The item “Other commitments in my life get in the way of studying for classes” had 2 missing responses. Therefore, all of the relative frequency percentages given in Table 4.2.2 below were calculated out of the total of 130 participants, with the exception of these items. That is, the percentage agreement and disagreement were calculated out of 128 for the items “I tell my professor when I don’t understand something from math class” and “Other commitments in my life get in the way of studying for classes”; and out of 129 for the items “The hardest part about solving word problems is understanding what is being asked”, “Thinking about the future I want makes me do more now to get that future”, and “I wait until right before a math test to start studying”.

In viewing the percentage of agreement and disagreement for each item of the SMPI across the entire sample, the majority of the responses for many of the items indicated the more desirable outcome on the item. That is, for most of the items that were positively coded (where higher levels of agreement indicate more desirable outcomes), more participants agreed than disagreed. Similarly, for the most of the negatively coded items (where lower levels of agreement, and thus, higher levels of disagreement, indicate more desirable outcomes), more participants disagreed than agreed. For example, for a positively coded item such as “My confidence in academic skills has increased this semester” 45.4% of participants agreed (here, agreement is associated with more desirable outcomes) and only 22.3% disagreed. Similarly, for a negatively coded item, for example, such as “I often play catch-up in my classes”, more people disagreed (41.5%), indicating a more desirable outcome, than agreed (37.7%). For 22 of the 33 items, it was the case more participants provided responses that aligned with the desirable outcomes than those that aligned with the less desirable outcomes.

For example, for each of the positively coded items of the FO scale, the vast majority of all participants agreed. For the item “I have a fairly clear idea of what I need to study now in order to have the career I want”, 70.8% of participants agreed and only 8.5% disagreed. For the item, “Thinking about the future I want makes me do more now to get that future” had an even greater percentage of agreement with 81.4% with only 3.9% disagreeing. Lastly, the item “I am taking the necessary steps to obtain the career I desire” had the highest levels of agreement of these items with 89.2% agreeing and only 1.5% disagreeing meaning that only 2 participants disagreed to this item. Thus, the levels of future orientation among all the participants in this study, according to the responses to these items, is quite high. Although this is a positive finding in general, and indicates that the participants perceive themselves as being strongly oriented to the future, this may call into question the validity and explanatory power of the FO scale. If the majority of all of the responses to these items indicate levels of agreement, the scale may not be accounting for enough variation in the responses of the participants (this potential limitation is discussed in Chapter 5).

For other items the difference in percentage agreement and disagreement was quite small. For example, the item “I am usually confident that I will do well on math tests” nearly as many people agreed to this item (34.6%) as disagreed (35.4%). This was also true for the items “I break down long-term assignments and/or class projects and work on them over time” (38.5% agreed and 37.7% disagreed) and “My grades in college math have influenced what degree I can pursue” (34.6% agreed and 31.5% disagreed). The responses indicating agreement and disagreement on each of these items responses account for nearly two thirds of all participants, and are split nearly evenly. This suggests that large subsets of participants differ greatly in their approach to mathematics assessments, their academic habits, and how their experiences in mathematics classes have impacted their future.

As just described, more than a third of all participants reported that their grades in college math have influenced what degree they can pursue. Nearly a third also agreed to the items “I’m just not good at math” (31.5% agreement), “Math and/or anything with numbers has been an obstacle to my academic success” (32.3% agreed), indicating that large proportion of individuals have limiting beliefs or experiences in mathematics. Even more notable is the fact that the majority of all participants overall agreed to the item “I have experienced difficulties in math since high school or before high school” (51.5% agreed and 40% disagreed), indicating that long-term difficulties in mathematics are more common than not among the surveyed participants. The item “The wording of math problems confuses me” also yielded agreement from the majority of all participants (53.1% agreed). The largest agreement, however, on a mathematics-related SMPI item was “The hardest part about solving word problems is understanding what is being asked” with 69% agreement. These results indicate that many participants share certain beliefs about the nature of mathematics, its problems, and themselves as learners.

The items just described indicate that the majority of all participants share some limiting beliefs, habits, and experiences in mathematics; however, majority sentiment on items indicating desirable outcomes in mathematics were also observed. For example, on the item “I push aside math assignments and do them last” the majority of all participants disagreed (51.5%), which indicates that the majority of participants do not avoid their mathematical work. However, it should be noted that a large proportion (28.5%) still agreed to this item, again indicating that a large number of individuals indeed take steps to avoid mathematical work. Over two-thirds (67.7%) of all participants agreed to the item “I can explain how I reach the correct answer on a math test”. The vast majority of all participants (82.3%) also agreed to the related item “When something confuses me, I think about it until I can make sense out of it." These items indicate that the majority of participants positively perceive their attention to detail, thought processes, and academic diligence.

In contrast, large proportions of participants negatively perceive their academic habits and outcomes, attention in class, and propensity to engage in class discussion. For the item "I find it hard to prioritize my time" 46.2% agreed and 29.2% disagreed. For the item, "Sometimes, I am disappointed in my test results because I studied a great deal" 61.5% agreed and 20.8% disagreed. The majority of all participants (50.8%) also agreed to the item "I often get distracted during class by feelings of boredom" and "I am hesitant to raise my hand in class even though I know the answer. In addition, more than two-thirds of participants (68.8%) agreed to the item "Other commitments in my life get in the way of studying for classes." Taken together, these results suggest that despite reported strength in other areas, many participants have negative perceptions of certain elements of their academic profile.

4.2.3 Analysis: SMPI Item Analysis Across Mathematics Course Level Subgroups

To determine whether at least one of the mathematics course level subgroups differed significantly from one another with respect to their responses to individual items of the SMPI, a set of Kruskal-Wallis tests were conducted. Recall, that the mathematics course level subgroups were formed after consulting university faculty and the university's course catalog to determine each course's prerequisite requirements; rigor and depth of the mathematical content; and whether these satisfy different subsequent quantitative major requirements. The subgroups of participants were students from introductory (commonly referred to as "remedial" or "developmental") courses in mathematics, which do not satisfy the university requirements in mathematics and serve as prerequisite courses for courses in the next level (e.g., introductory/intermediate algebra); a mathematics course that satisfies the general education requirement for students not in the STEM (science, technology, engineering, or mathematics) majors (e.g., elementary statistics, mathematics for elementary education, etc.); a required mathematics course for STEM majors (e.g., calculus I, calculus II, intermediate statistics, etc.); and a major course in mathematics (e.g., foundations of mathematics, discrete mathematics, real analysis, etc.). Recall that, respectively, according to the order just described, the mathematics course level subgroups will henceforth be referred to as the "developmental level", the "general level", the "STEM level", and the "mathematics level". In total, there were 11 participants in the developmental subgroup, 55 in the general level subgroup, 44 in the STEM level subgroup, and 20 in the mathematics level subgroup.

Table 4.2.3 below provides the information for the Kruskal-Wallis tests conducted on individual items of the SMPI across the mathematics course level subgroups. The individual item is given, as is the H-statistic for each of the tests (all calculated with degrees of freedom 3) and the corresponding p-value. Asterisks are marked on the p-values to indicate levels of significance; single asterisk (*) indicate statistical significance at the $\alpha = .05$ level, and double asterisk (**) indicate statistical significance at the $\alpha = .01$ level. Only the Kruskal-Wallis tests that were significant are detailed in this table; all other analysis on individual items were not significant. It should be noted that a statistically significant difference was found between the mathematics course level subgroups on each of the items comprising the Limiting Mathematics Beliefs, Habits, and Experiences in Mathematics (LBHEM) scale. Each of these differences were significant at the $\alpha = .01$ level.

Table 4.2.3

Significant Kruskal-Wallis Tests on Individual SMPI Items Across Mathematics Course Level Subgroups

SMPI Item	H(3)	*p*
I'm just not good at math.	17.146	<.001**
The wording of math problems confuses me.	20.037	<.001**
I tell my professor when I don't understand something from math class.	9.902	.019*
Math and/or anything with numbers has been an obstacle to my academic success.	18.993	<.001**
I expect to learn the math I have learned in my future career.	34.586	<.001**
I have experienced difficulties in math since high school or before high school.	18.884	<.001**
I am usually confident that I will do well on math tests.	20.425	<.001**
I push aside math assignments and do them last.	12.737	.005**
I can explain how I reach the correct answer on a math test.	19.675	<.001**
Sometimes, I am disappointed in my test results because I studied a great deal.	9.224	.026*
I am hesitant to raise my hand in class even though I know the answer.	15.984	.001**
I fear that if I ask for help, my professor will think less of me.	15.540	.001*
I often get distracted during class by feelings of boredom.	11.958	.008**

Note. The H-test statistic has been adjusted for ties. Also, the degrees of freedom for each test is 3.

These findings indicate that participants enrolled in different mathematics course levels expressed differing levels of limiting characteristics related to mathematics. Additionally, statistically significant differences were observed on items of the HEF factor, an item indicating feelings of boredom in class, an item indicating the expectation of using mathematics in participants' future careers, and disappointment with assessment scores despite studying.

Table 4.2.4 provides the significant pairwise comparisons between individual mathematics course level subgroups on each of the SMPI items for which the Kruskal-Wallis test was significant. Recall that for these pairwise tests (i.e., utilizing Dunn's test) the difference in mean rank is calculated, which is then used to calculate a standardized test-statistic which yields the significance value. Therefore, in the table, the individual item for which the original Kruskal-Wallis test was significant is given, followed by the two subgroups under analysis, the difference in their mean ranks, the corresponding standardized test statistic, and corresponding p-value. Since each analysis in the table is pairwise, the two groups in each analysis are given, indicated by "Group A" and "Group B" so that it is clear how the difference in mean ranks was calculated. For example, if "Group A" is indicated to be the mathematics level subgroup and "Group B" is indicated to be the general level subgroup, the difference in mean ranks, indicated by (A−B) provides the difference in mean ranks calculated by subtracting the mean rank of the general level subgroup from the mean rank of the mathematics level subgroup.

Listing the differences in this way helps the reader view which mean ranks were larger or smaller for a given variable. Recall that, if the mean rank in this fourth column is negative, then the mean rank for the first group exceeded that of the second group. If the mean rank of the second group exceeded that of the first group, the value in the fourth column indicated by (A−B) would be positive. Also, only the pairwise analyses that were significant using the Bonferroni correction for multiple tests are listed in the table. Asterisks are marked on the p-values to indicate levels of significance; single asterisk (*) indicate statistical significance at the $\alpha = .05$ level, and double asterisk (**) indicate statistical significance at the $\alpha = .01$ level. Recall that although an individual pairwise test may be significant, the same test may not be significant when the significance value is adjusted for the influence of multiple tests (i.e., adjusted using the Bonferroni correction). To be conservative in the reporting of findings, only pairwise analyses that were significant with the Bonferroni correction are reported in the table below. Also, the mathematics course level subgroups have been abbreviated to "developmental," "general," "STEM," and "mathematics".

Table 4.2.4

Significant Pairwise Comparisons Between Mathematics Course Level Subgroups on Individual SMPI Items

Item	Group A	Group B	Mean Rank Diff. (A-B)	Std. Test Stat.	Sig. (Adj.)
I'm just not good at math.	Mathematics	General	-37.389	-3.910	.001**
	Mathematics	Developmental	-38.998	-2.837	.027*
The wording of math problems confuses me.	Mathematics	STEM	32.552	3.342	.005
	Mathematics	General	-41.523	-4.402	<.001**
	Mathematics	Developmental	-39.768	-2.933	.020*
I tell my professor when I don't understand something from math class.	General	Mathematics	-26.605	-2.859	.026*
	STEM	Mathematics	-26.208	-2.717	.040*
Math and/or anything with numbers has been an obstacle to my academic success.	Mathematics	General	-36.533	-3.824	.001**
	Mathematics	Developmental	-36.650	-2.668	.046*
	STEM	General	-21.484	-2.903	.022*
I expect to use the math I have learned in my future career.	Developmental	STEM	-41.659	-3.397	.004**
	Developmental	Mathematics	-69.595	-5.096	<.001**
	General	Mathematics	-45.395	-4.779	<.001**
	STEM	Mathematics	-27.936	-2.847	.026*
I have experienced difficulties in math since high school or before high school.	Mathematics	General	-37.639	-3.930	.001**
	STEM	General	-21.980	-2.963	.018*
I am usually confident that I will do well on math tests.	Developmental	Mathematics	-49.359	-3.617	.002**
	General	Mathematics	-39.605	-4.173	<.001**
	STEM	Mathematics	-32.552	-3.321	.005**
I push aside math assignments and do them last.	Mathematics	General	-31.641	-3.351	.005**
I can explain how I reach the correct answer on a math test.	Developmental	Mathematics	-48.120	-3.653	.002**
	General	Mathematics	-34.057	-3.717	.001**
Sometimes, I am disappointed in my test results because I studied a great deal.	STEM	General	-20.186	-2.748	.036*
I am hesitant to raise my hand in class even though I know the answer	Mathematics	STEM	-30.666	-3.095	.012*
	Mathematics	General	-33.834	-3.527	.003**
I fear that if I ask for help, my professor will think less of me.	Developmental	General	-39.145	-3.268	.007**
	Mathematics	General	-26.773	-2.827	.028*
I often get distracted during class by feelings of boredom.	Mathematics	General	-30.941	-3.230	.007**

In viewing first the pairwise analyses on the items of the LBHEM scale (recall, that the items of this scale are “I’m just not good at math”, “The wording of math problems confuses me”, “Math and/or anything with numbers has been an obstacle to my academic success”, “I have experienced difficulties in math since high school or before high school”, “I am usually confident that I will do well on math tests”, “I push aside math assignments and do them last,” and “I can explain how I reach the correct answer on a math test”), participants in successively higher mathematics course levels had more desirable responses in general than those in lower course levels. That is, the participants in the mathematics course level subgroup had the most desirable responses on all of the mathematics-related items. On positively coded items (which indicates that higher levels of agreement are associated with more desirable outcomes), they had the highest mean ranks (indicating a greater proportion of their responses were aligned with the desirable outcome) than all of the other groups; similarly, on negatively coded items, they had the lowest mean ranks or the lowest levels of agreement.

In each of the pairwise analyses on the LBHEM items that were significant, the mathematics course level subgroup had significantly more desirable scores than at least one of the other mathematics course level subgroups. Moreover, in every significant analysis of this kind, the mathematics course level subgroup had significantly more desirable responses than the general level group. In these analyses where significant differences were found between other groups than just with the mathematics course level subgroup (e.g., between the developmental, general, and STEM level subgroups) the subgroups that were more advanced had the more desirable outcomes. That is, for pairwise analyses on LBHEM items that were significant between these groups, the STEM level had more desirable responses than those in the general and developmental levels, and the general level had more desirable responses than those in the developmental level.

Taken together, the results suggest a hierarchy of desirable responses that aligns with the hierarchy of the mathematics course level subgroups. In other words, those in the higher mathematics course level subgroups had more desirable responses on each of the mathematics-related items than their peers in lower level course level subgroups, with the mathematics level subgroup having the most desirable responses overall. Moreover, since these items were part of the LBHEM scale and measured limiting characteristics related to mathematics, the participants in the higher level mathematics course level subgroups reported fewer limiting characteristics than their peers in the lower levels, with the mathematics course level subgroup reporting the least limiting characteristics overall.

On several of the items not on the LBHEM factor, participants in the mathematics course level subgroup also had more desirable responses overall. The responses of the participants in the mathematics course level subgroup were significantly different than each of the other subgroups for the item "I expect to use the math I have learned in my future career". Participants in the mathematics course level subgroup also had significantly more desirable responses than those in the STEM and general level subgroups on two of the items from the Hesitancy to Engage with Faculty (HEF) scale, which were "I am hesitant to raise my hand in class even though I know the answer" and "I fear if I ask for help, my professor will think less of me." This finding suggests that mathematics level students are not hesitant to speak up in class and ask for help.

Additionally, the latter of these two items, the developmental students had significantly more desirable responses than those in the general level. One interpretation of this may be that participants in the developmental level are in greater need of help on mathematical work and/or they are in supportive developmental classes with faculty who are understanding of these needs.

Participants from the mathematics course level subgroups also had more desirable responses than those in the general level group on the item "I often get distracted during class by feelings of boredom". Finally, for the item "Sometimes, I am disappointed in my test results because I studied a great deal", participants in the STEM group had significantly more desirable responses than those in the general level.

4.2.4 Analysis: SMPI Item Analysis Across Other Subgroups

Gender Subgroups: Males and Females To determine whether males and females differed significantly each with respect to their responses to individual items of the SMPI, a set of Mann-Whitney U-Tests were conducted on each of the SMPI items. Table 4.2.5 provides the results of each of the Mann-Whitney U-Tests that were significant (all other item analyses between the two groups were not significant). The table includes the individual item under analysis, the mean rank of both the male and female subgroups, the U-test statistic, the corresponding z-test statistic, and the corresponding significance level. Asterisks are again marked on the p-values to indicate levels of significance; a single asterisk (*) indicate statistical significance at the $\alpha = .05$ level, and a double asterisk (**) indicate statistical significance at the $\alpha = .01$ level.

Since each of the analyses given in the table were significant (meaning a significant difference between the two groups with respect to their responses on the variable was found), the mean ranks for each group must be consulted to determine which subgroup had higher levels of agreement or disagreement to each item. Moreover, the mean ranks provide an indication of which group had more desirable scores on each respective item. Also, if the z-test statistic in the fourth column is negative, then the mean rank for the males exceeded that of the females. Alternatively, if the mean rank of the females exceeded that of the males, the z-test statistic in the fourth would be positive. For positively-coded items, higher mean ranks indicate more desirable responses, and, conversely, for negatively-coded items lower mean ranks indicate more desirable responses.

Table 4.2.5

Significant Mann-Whitney U-Tests Between Males and Females on Individual SMPI Items

Item	Mean Rank: Males	Mean Rank: Females	U-Test Statistic	Z-Test Statistic	Sig.
I use my time between classes productively.	51.22	72.03	1285	-3.374	<.001**
I break down long-term assignments and/or class projects and work on them over time.	52.57	71.18	1351	-2.864	.004**
I often play catch up in my classes	74.91	57.15	1376.5	2.728	.006**
I am usually confident that I will do well on math tests.	71.91	59.03	1523.5	1.989	.047*
I give myself enough time to really read course materials	55.95	69.06	1516.5	-2.062	.039*
I am taking the necessary steps to obtain the career I desire.	54.99	69.66	1469.5	-2.498	.013*
I study regularly to be successful in college.	55.13	69.57	1476.5	-2.261	.024*
Other commitments in my life get in the way of studying for classes.	71.89	58.16	1475.5	2.166	.030*
I am doing better than I thought I would in college.	55.19	69.53	1479.5	-2.220	.026*

With the exception of two items ("I am usually confident that I will do well on math tests" and "Other commitments in my life get in the way of studying for classes") females had more desirable responses on all of the items where the difference between the two groups was significant. Based upon the results above, it seems as if the female participants have more desirable academic habits and levels future orientation than males. The results indicate that females report better time management, study skills, and are better oriented to the future. Additionally, they report viewing their performance in college more positively.

Males, however, report being more confident that they will do well on mathematics assessments. Despite not viewing their overall performance in college more positively than females, they seem to report a more positive view of themselves within the mathematical domain. Also, males also reported that other commitments in their life get in the way of studying for classes at a rate greater than that of females. Therefore, taken together, and with the exception of how confidently males report approaching mathematics assessments, females reported having more desirable outcomes overall on items related to academic habits of mind and the influence of extracurricular commitments.

Academic Major Subgroups To determine whether at least one of the academic major subgroups differed significantly from one another with respect to their responses to individual items of the SMPI, a set of Kruskal-Wallis tests were conducted. Recall, that six academic major subgroupings were formed based upon responses to the demographic subgroup questions at the beginning of the SMPI. The six subgroupings (see Table 4.2.1 for a breakdown of the size of each group) are: business-related majors (e.g., business administration, business in finance, business management, etc.), education-related majors (e.g., early childhood education, elementary education, collaborative education, special education, etc.), health and human services-related majors (e.g., public health, nursing, healthcare studies, psychology, exercise science, social work, etc.), mathematics-related majors (e.g., mathematics, applied mathematics, mathematics with secondary education certification, etc.), science-related majors (e.g., physics, computer science, biochemistry, biology, etc.), and students with no majors (e.g., undeclared, undecided, etc.).

The results for the Kruskal-Wallis tests on each individual item are given in Table 4.2.6. The individual item is given, as is the H-statistic for each of the tests (all calculated with degrees of freedom 5) and the corresponding p-value. Asterisks are marked on the p-values to indicate levels of significance; single asterisk (**) indicate statistical significance at the $\alpha = .05$ level, and double asterisk (**) indicate statistical significance at the $\alpha = .01$ level. Only the Kruskal-Wallis tests that were significant are detailed in this table; all other analysis on individual items were not significant.

Table 4.2.6

Significant Kruskal-Wallis Tests on Individual SMPI Items Across Academic Majors

SMPI Item	*H(5)*	*p*
I'm just not good at math.	18.031	.003**
The wording of math problems confuses me.	18.973	.002**
Math and/or anything with numbers has been an obstacle to my academic success.	21.043	<.001**
I expect to use the math I have learned in my future career.	34.005	<.001**
I have experienced difficulties in math since high school or before high school.	15.924	.007**
I am usually confident that I will do well on math tests.	18.358	.003**
I push aside math assignments and do them last.	19.205	.002**
I can explain how I reach the correct answer on a math test.	16.165	.006**
Sometimes, I am disappointed in my test results because I studied a great deal.	11.739	.039*
I fear that if I ask for help, my professor will think less of me.	17.874	.003**
I often get distracted during class by feelings of boredom.	14.132	.015*

Note. The H-test statistic has been adjusted for ties. Also, the degrees of freedom for each test is 5.

The next table, Table 4.2.7, provides the significant pairwise comparisons between individual academic major subgroups on each of the SMPI items for which the Kruskal-Wallis test was significant. Following the format and structure of Table 4.2.5 described in the previous section, the individual item for which the original Kruskal-Wallis test was significant is given, followed by the two subgroups under analysis, the difference in their mean ranks, the corresponding standardized test statistic, and corresponding p-value. Recall again that, if the mean rank in this fourth column is negative, then the mean rank for the first group exceeded that of the second group. If the mean rank of the second group exceeded that of the first group, the value in the fourth column indicated by (A−B) would be positive.

Also, only the pairwise analyses (utilizing Dunn's test) that were significant using the

Bonferroni correction for multiple tests are listed in the table. Recall that although an individual pairwise test may be significant, the same test may not be significant when the significance value is adjusted for the influence of multiple tests (i.e., adjusted using the Bonferroni correction). To be conservative in the reporting of findings, only pairwise analyses that were significant with the Bonferroni correction are reported in the table below. The academic major subgroupings have been abbreviated to “Business”, “Science”, “HHS” (for majors in Health and Human Services), “Education”, and “Mathematics”.

In viewing the results, significant differences were found between individual academic major subgroups on each of the items of the LBHEM scale (recall, that the items of this scale are “I’m just not good at math”, “The wording of math problems confuses me”, “Math and/or anything with numbers has been an obstacle to my academic success”, “I have experienced difficulties in math since high school or before high school”, “I am usually confident that I will do well on math tests”, “I push aside math assignments and do them last,” and “I can explain how I reach the correct answer on a math test”). On each of these, the participants in the mathematics major had significantly more desirable results than at least one of the other academic subgroupings. One consistent pattern that emerged was that participants in the mathematics major had significantly more desirable responses than participants in HHS majors on each of the LBHEM items. Additionally, mathematics majors reported more desirable responses than those in the science majors on the LBHEM items of “The wording of math problems confuses me”, “I am usually confident that I will do well on math tests”, and “I push aside math assignments and do them last”.

Table 4.2.7

Significant Pairwise Comparisons Between Academic Major Subgroups on Individual SMPI Items

Item	Group A	Group B	Mean Rank Diff. (A-B)	Std. Test Stat.	Sig. (Adj.)
I'm just not good at math.	Mathematics	HHS	-40.721	-4.087	.001**
The wording of math problems confuses me.	Mathematics	Science	-35.033	-3.754	.003**
	Mathematics	HHS	-39.996	-4.070	.001**
Math and/or anything with numbers has been an obstacle to my academic success.	Mathematics	Education	-37.395	-3.052	.034*
	Mathematics	HHS	-43.567	-4.376	<.001**
I expect to use the math I have learned in my future career.	Business	Mathematics	-70.282	-3.899	.001**
	HHS	Mathematics	-53.039	-5.358	<.001**
	Science	Mathematics	-36.905	-3.927	.001**
I have experienced difficulties in math since high school or before high school.	Mathematics	HHS	-36.758	-3.684	.003**
I am usually confident that I will do well on math tests.	Business	Mathematics	-54.714	-3.038	.036*
	HHS	Mathematics	-36.928	-3.734	.003**
	Science	Mathematics	-28.731	-3.060	.033*
I push aside math assignments and do them last.	Mathematics	Science	-29.600	-2.168	.023*
	Mathematics	HHS	-38.642	-3.927	.001**
I can explain how I reach the correct answer on a math test.	HHS	Mathematics	-35.697	-3.739	.003**
Sometimes, I am disappointed in my test results because I studied a great deal.	Science	HHS	-25.154	-3.102	.029*
I fear that if I ask for help, my professor will think less of me.	Business	HHS	-57.929	-3.341	.013*
I often get distracted during class by feelings of boredom.	Mathematics	Science	-28.260	-2.982	.043*
	Mathematics	HHS	-33.181	-3.324	.013*

Note. HHS stands for health and human services

This was also observed on the items "I expect to use the math I have learned in my future career" and "I often get distracted during class by feelings of boredom" where mathematics majors also had significantly higher responses than those in the HHS majors. On the two items for which this was not the case, a significant difference between the mathematics major subgroup and other subgroups were not found. The first of these items was "Sometimes, I am disappointed in my test results because I studied a great deal", where participants in the science majors had significantly more desirable responses than those in the HHS majors. The other item was "I fear that if I ask for help, my professor will think less of me", where students in the Business majors had significantly more desirable responses than those in the HHS major.

Academic Year Subgroups To determine whether at least one of the academic year subgroups differed significantly from one another with respect to their responses to individual items of the SMPI, a set of Kruskal-Wallis tests were conducted. Recall that the academic year subgroups were first-year ($n = 46$), sophomore ($n = 26$), junior ($n = 34$), and senior ($n = 22$). There were only two items for which a significant difference between individual academic year subgroups was found. Both items were from the Hesitancy to Engage with Faculty (HEF) scale. A significant difference was found between the responses of participants from different academic year subgroups on the item "I am hesitant to raise my hand in class even though I know the answer" [$H(3) = 8.570,\ p = .036$]. Participants in the senior year had significantly more desirable responses (i.e., lower levels of agreement on this item) than first year students (mean rank difference: $25.114, z = 2.677, p = .045$).

Significant differences were also found between the responses of participants from different academic year subgroups on the item "I fear that if I ask for help, my professor will think less of me" [$H(3) = 9.740,\ p = .021$] (note that in both of these analyses, the H-statistic was adjusted for ties). Pairwise post hoc analyses utilizing Dunn's test with a Bonferroni correction revealed that participants in the senior year subgroup had significantly more desirable responses to this item than both first- year participants (mean rank difference: $24.505, z = 2.643, p = .049$) and sophomore participants (mean rank difference: $30.267, z = 2.921, p = .021$). One possible interpretation for these findings is that students become less hesitant to engage with faculty or have less negative feelings about doing so the longer they are in the college environment.

Mathematics Courses Taken in Senior Year of High School To determine whether at least one of the senior year high school mathematics course subgroups differed significantly from one another with respect to their responses to individual items of the SMPI, a set of Kruskal-Wallis tests were conducted. Recall, that these subgroups were formed according to the level and type of mathematics courses participants reported taking during their senior year of high school. The subgroups were: participants that did not take a math class their senior year of high school; those who took courses related to algebra (including Algebra, Algebra II, Algebra III, Algebra and Statistics, ALEKS Program, College Algebra, Introduction to College Mathematics, Pre-Algebra, and Trigonometry and Algebra); those who took general courses (including accounting, general statistics, basic statistics, probability, and business mathematics); those who took courses in precalculus (including precalculus, advanced algebra and precalculus, precalculus and trigonometry, and precalculus and statistics); and those who took advanced courses (including Calculus, AP Calculus, AP Statistics, and Multivariable Calculus). To be listed in the following tables, these five subgroupings were abbreviated, respectively, as “None”, “Algebra”, “General”, “Precalculus”, and “Advanced”.

The results for the Kruskal-Wallis tests on each individual item are given in Table 4.2.8. The individual item is given, as is the H-statistic for each of the tests (all calculated with degrees of freedom 5) and the corresponding p-value. Asterisks are marked on the p-values to indicate levels of significance; single asterisk (*) indicate statistical significance at the $\alpha = .05$ level, and double asterisk (**) indicate statistical significance at the $\alpha = .01$ level. Only the Kruskal-Wallis tests that were significant are detailed in this table; all other analysis on individual items were not significant.

Table 4.2.8

Significant Kruskal-Wallis Tests on Individual SMPI Items Across Senior Year High School Mathematics Course Subgroups

SMPI Item	H(4)	*p*
I’m just not good at math	18.123	.001
Math and/or anything with numbers has been an obstacle to my academic success.	20.418	<.001
I have experienced difficulties in math since high school or before high school.	25.863	<.001
I am usually confident that I will do well on math tests.	10.679	.030
I wait until right before a math test to start studying.	10.962	.027
I push aside math assignments and do them last.	12.107	.017
I can explain how I reach the correct answer on a math test.	11.690	.020
Sometimes, I am disappointed by my test results because I studied a great deal.	14.633	.006

Note. The H-test statistic has been adjusted for ties. Also, the degrees of freedom for each test is 3.

The next table, Table 4.2.9, provides the significant pairwise comparisons between individual academic major subgroups on each of the SMPI items for which the Kruskal-Wallis test was significant. Following the format and structure of Table 4.2.4 described in the previous section, the individual item for which the original Kruskal-Wallis test was significant is given, followed by the two subgroups under analysis, the difference in their mean ranks, the corresponding standardized test statistic, and corresponding p-value. Also, only the pairwise analyses (utilizing Dunn’s test) that were significant using the Bonferroni correction for multiple tests are listed in the table. Recall again that, if the mean rank in this fourth column is negative, then the mean rank for the first group exceeded that of the second group. If the mean rank of the second group exceeded that of the first group, the value in the fourth column indicated by (A−B) would be positive. Also, recall that although an individual pairwise test may be significant, the same test may not be significant when the significance value is adjusted for the influence of multiple tests (i.e., adjusted using the Bonferroni correction). To be conservative in the reporting of findings, only pairwise analyses that were significant with the Bonferroni correction are reported in the table below. The academic major subgroupings have been abbreviated to "Business", "Science", "HHS" (for majors in Health and Human Services), "Education", and "Mathematics". Refer to Table 4.2.1 for a breakdown of the sizes of each group.

Table 4.2.9

Significant Pairwise Comparisons Between Senior Year High School Mathematics Course Subgroups on Individual SMPI Items

Item	Group A	Group B	Mean Rank Diff. (A-B)	Std. Test Stat.	Sig. (Adj.)
I'm just not good at math.	Advanced	General	25.847	2.87	.041
	Advanced	Precalculus	26.660	3.16	.016
	Advanced	Algebra	31.767	3.31	.009
	Advanced	None	37.294	3.18	.015
Math and/or anything with numbers has been an obstacle to my academic success.	Advanced	General	31.417	3.483	.005
	Advanced	Algebra	31.942	3.32	.009
	Advanced	Precalculus	32.310	3.81	.001
I have experienced difficulties in math since high school or before high school.	Advanced	General	31.233	3.464	.005
	Advanced	Precalculus	34.905	4.125	<.001
	Advanced	Algebra	42.083	4.378	<.001
I am usually confident that I will do well on math tests.	Advanced	Algebra	-29.858	-3.128	.018
I wait until right before a math test to start studying.	Advanced	Algebra	-30.972	-3.226	.013
I push aside math assignments and do them last.	Advanced	Algebra	30.100	3.169	.015
I can explain how I reach the correct answer on a math test.	Advanced	Algebra	-30.083	-3.264	.011
Sometimes, I am disappointed by my test results because I studied a great deal.	Advanced	Algebra	31.433	3.303	.010

Significant differences were found between individual subgroups on each of the items of the LBHEM scale (recall, that the items of this scale are "I'm just not good at math", "The wording of math problems confuses me", "Math and/or anything with numbers has been an obstacle to my academic success", "I have experienced difficulties in math since high school or before high school", "I am usually confident that I will do well on math tests", "I push aside math assignments and do them last," and "I can explain how I reach the correct answer on a math test"). On each of these, the participants in the advanced senior year high school mathematics course subgroup had significantly more desirable results than at least one of the other subgroupings. One consistent pattern that emerged was that participants in the advanced subgroup had significantly more desirable responses than the algebra group on each of the LBHEM items. On the item "I'm just not good at math" the advanced level group had significantly more desirable responses (i.e., lower levels of agreement) than each of the other subgroups. On the only item that was not on the LBHEM scale ("Sometimes, I am disappointed by my test results because I studied a great deal"), participants in the advanced level subgroup again had significantly more desirable responses than those in the algebra level subgroup.

Taken together, these results indicate that those who took advanced level courses their senior year of high school reported fewer limiting characteristics related to mathematics than their peers, particularly their peers who took algebra level courses their senior year. Moreover, advanced level students had more desirable characteristics overall with respect to items related to mathematics. Additionally, participants who took advanced level courses their senior year of high school also belong to different mathematics course level subgroups (in particular, 4 belong to the general level, 15 belong to the STEM level, and 11 belong to the mathematics course level subgroups). Therefore, these results provide information about how high school mathematics course enrollment, separate from collegiate mathematics course enrollment, is related to limiting characteristics in mathematics.

4.2.5 Analysis: SMPI Scale Analysis

Recall that the scales of the SMPI were derived from a principal axis factoring analysis with a Promax rotation on all SMPI items (refer to Chapter 3 for the details of this analysis). Five reliable scales emerged, each of which comprise items that were included to measure certain psychoeducational facets of college students' educational profile. These scales are the Limiting Beliefs, Habits, and Experiences Related to Mathematics Scale (LBHEM); the Academic Habits of Mind Scale (AHM); the Future Orientation Scale (FO); the Academic Self-Concept (ASC) Scale; and the Hesitancy to Engage with Faculty (HEF) scale. Recall that all of the items within each individual scale were coded on the same 5-point Likert scale. However, for some items, lower values of the scale (1=strongly disagree, 2=agree) may algin with higher values on another item (4=agree, 5=strongly agree). For example, in the LBHEM factor, higher values for the item "I'm just not good at math" are associated with lower values for the item "I am usually confident I will do well on math tests" which each indicate less desirable or limiting outcomes. Therefore, in the development of the scales described below, some items needed to be reverse coded so that each item aligns with one another on the same measured 5-point scale. The scales described above are each included in Table 4.2.10 below. The name of each scale and its corresponding calculated reliability coefficient (Cronbach's α) are given alongside the items that comprise each scale.

Recall that composite scores were created from the scales described above for the purposes of analyzing responses related to the psychoeducational facets described by each factor/scale. The composites were created by directly adding the responses for the items on each scale. Recall that each of the scales were coded in the same structure (some items needed to be reverse coded before inclusion on the scale). The composite score for the LBHEM scale, referred to as LBHEM Composite, consisted of 7 items, when directly added resulted in a minimum composite score of 8 and a maximum of 35. The mean of the LBHEM Composite was 20.09, the median was 20, and the standard deviation was 6.2.

Table 4.2.10

Scales Derived from the SMPI

Scale and Reliability Coefficient	Items of the Scale
Limiting Beliefs, Habits, and Experiences in Mathematics (LBHEM) Reliability: $\alpha = .860$	I'm just not good at math.
	Math and/or anything with numbers has been an obstacle to my academic success.
	I have experienced difficulties in math since high school or before high school.
	I am usually confident that I will do well on math tests. (Reverse coded)
	I can explain how I reach the correct answer on a math test. (Reverse coded)
	I push aside math assignments and do them last.
	The wording of math problems confuses me.
Academic Habits of Mind (AHM) Reliability: $\alpha = .742$	I often play catch-up in my classes. (Reverse coded)
	I wait until right before a math test to start studying. (Reverse coded)
	I give myself enough time to really read course materials.
	I settle for just passing my courses. (Reverse coded)
	I break down long-term assignments and work on them over time.
	I study regularly to be successful in college.
	I find it hard to prioritize my time. (Reverse coded)
Future Orientation (FO) Reliability: $\alpha = .637$	I am taking the necessary steps to obtain the career I desire.
	I have a fairly clear idea of what I need to study now in order to have the career I want.
	Thinking about the future I want makes me do more now to get that future.
Academic Habits of Mind (ASC) Reliability: $\alpha = .635$	My confidence in academic skills has increased this semester.
	I am doing better than I thought I would in college.

The AHM Composite was calculated by summing the 7 AHM Scale items, and ranged from 9 to 34, with a mean of 22.93, a median of 22, and a standard deviation of 4.9.The FO Composite was calculated by summing the 3 FO Scale items, and ranged from 5 to 15, with a mean of 11.99, a median of 12, and a standard deviation of 1.84. The ASC Composite was calculated by summing the 2 ASC Scale items, and ranged from 2 to 19, with a mean of 6.67, a median of 7, and a standard deviation of 1.844. Lastly, the HEF Composite was calculated by summing the 3 HEF Scale items and ranged from 3 to 15, with a mean of 8.07, a median of 8, and a standard deviation of 3.153.

4.2.6 Analysis: SMPI Scale Analysis Across Mathematics Course Level Subgroups

Limiting Beliefs, Habits, and Experiences Related to Mathematics (LBHEM) Scale

As mentioned in the previous section, the analysis of individual SMPI items across the mathematics level subgroups suggests a hierarchy of desirable responses that aligns with the hierarchy of the mathematics course level subgroups. In other words, those in the higher mathematics course level subgroups had more desirable responses on each of the mathematics-related items than their peers in lower level course level subgroups, with the mathematics level subgroup having the most desirable responses overall. The inferential analyses in the previous section (Kruskal-Wallis H-Tests with post hoc analyses using Dunn’s test) align with the descriptive makeup of the LBHEM factor.

Aligning with the analyses from the previous section, the responses indicate that the mathematics course level subgroup tend to have the most desirable responses overall, followed respectively by the STEM, general, and developmental subgroups. This can be seen by viewing the percentage agreement to positively coded items (where higher levels of agreement are associated with more desirable outcomes) and percentage disagreement to negatively coded items (where lower levels of agreement, i.e., higher levels of disagreement correspond to more desirable outcomes). Figures 4.2.11 and 4.2.12 depicts descriptive breakdown of each item of the LBHEM factor across the mathematics course level subgroups.

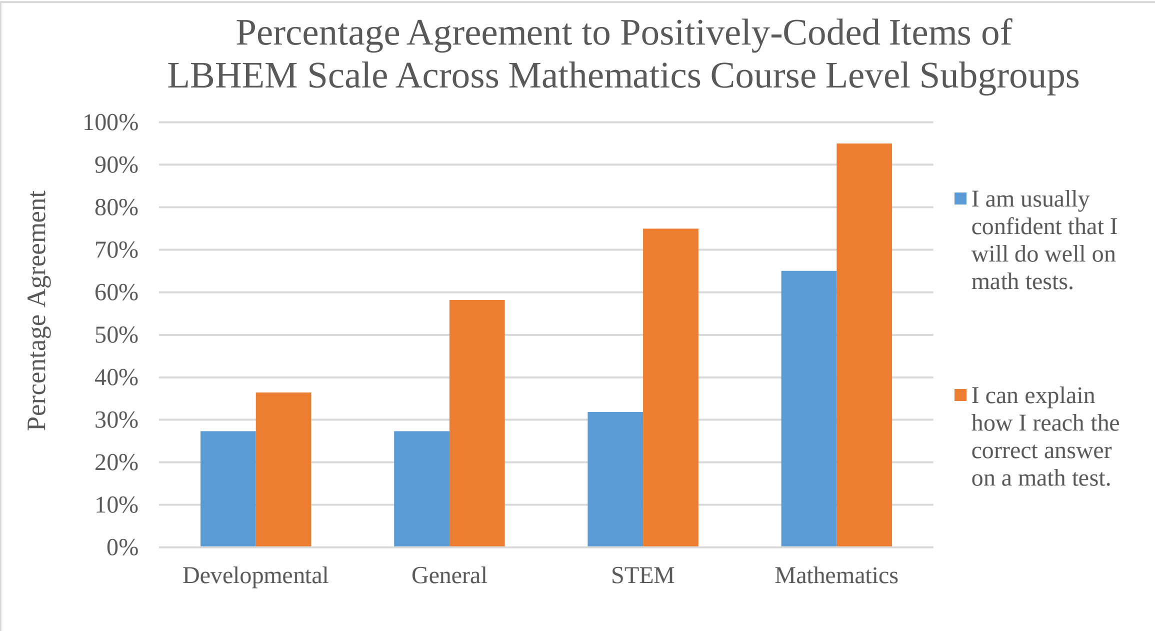

Figure 4.2.11: Graph (Sid-By-Side Bar Graph) Depicting Percentage Agreement to Positively-Coded LBHEM Scale Items Across Mathematics Course Level Subgroups

The percentage agreement and disagreement is given for each LBHEM item across each of the mathematics course level subgroups. Figure 4.2.11 depicts the percentage agreement to the positively-coded items of the LBHEM Scale across each of the mathematics course level subgroups. Figure 4.2.12 depicts the percentage of agreement to the negatively-coded items of the LBHEM Scale across each of the mathematics course level subgroups. The bars in each graph indicate the relative frequency percentage of each type of response (agreement or disagreement) for each of the respective mathematics course level subgroups. The data tables used to make these graphs (which include the percentage disagreement to each item) are given in Appendix C (Table 4.2.11 and Table 4.2.12).

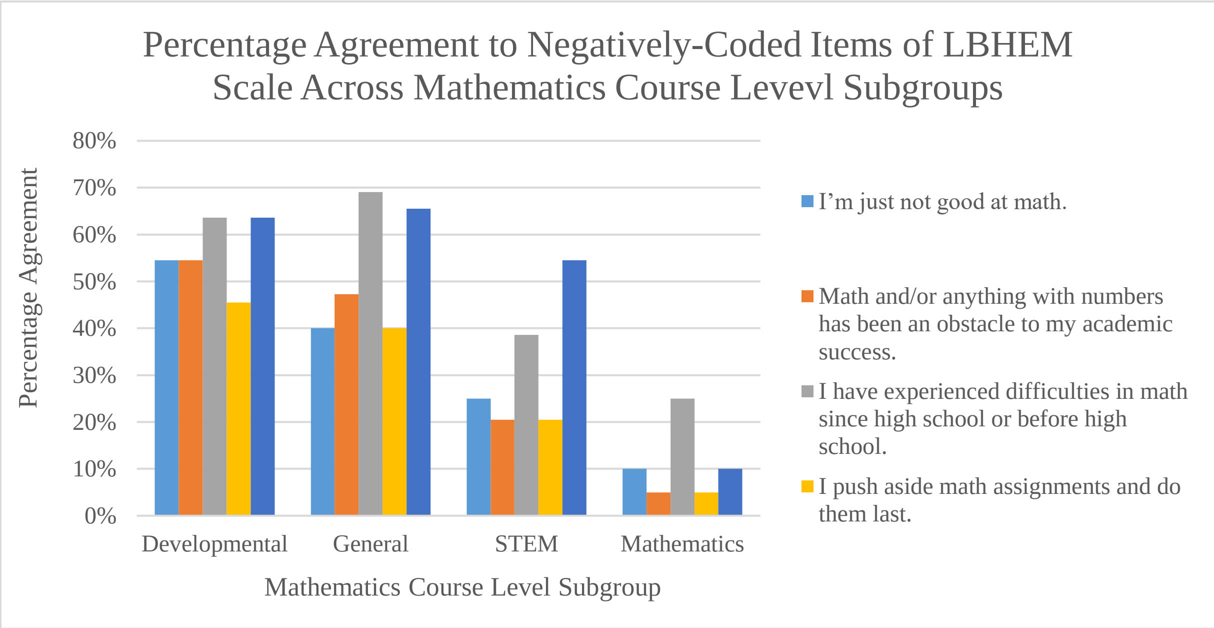

Figure 4.2.12: Graph (Sid-By-Side Bar Graph) Depicting Percentage Agreement to Negatively-Coded LBHEM Scale Items Across Mathematics Course Level Subgroups

As mentioned above, the responses indicate that the mathematics course level subgroup tend to have the most desirable responses overall, followed respectively by the STEM, general, and developmental subgroups. For example, for the item "I am usually confident that I will do well on math tests" (which is positively coded, indicating higher levels of agreement are more desirable), 27.3% of the participants in the developmental and general groups either agreed or strongly agreed, followed by 31.8% of the STEM group, and 65% of the mathematics group. Conversely, disagreement to this item also indicates less desirable outcomes. On this item, 54.5% of the developmental group participants disagreed, followed by 47.3% of the general group, 31.8% of the STEM group, and 0% of the mathematics group. An example involving a negatively coded item is the item "Math and/or anything with numbers has been an obstacle to my academic success" (where higher levels of disagreement is associated with more desirable outcomes). For this item, 36.4% disagreed from the developmental group, followed by 38.2% from the general group, 59.1% from the STEM group, and 90% from the mathematics group.

Again, conversely, agreement to this item indicates less desirable outcomes, and such responses were observed less frequently across the different course level subgroups. For this item, 54.5% of the developmental participants agreed to this item, followed by 47.3% of general level participants, 20.5% of STEM, and only 5% of the mathematics level participants.

A similar pattern can also be viewed through the descriptive statistics of the LBHEM Scale Composite Score. Recall from above that the composite score was obtained by summing the responses to each item of the LBHEM scale. Table 4.2.13 provides the descriptive statistics of the LBHEM Composite Score.

The LBHEM Scale Composite Score, like the LBHEM factor itself, provides a summative measure of the limiting characteristics related to mathematics reported by participants on the SMPI.

Table 4.2.13

Descriptive Statistics of LBHEM Scale Composite Score Across Entire Sample and Mathematics Course Level Subgroups

Subgroup	LBHEM Scale Composite Score		
	Mean	Median	Standard Deviation
All Participants	20.09	20	6.2
Developmental Course Level	23	25	7.13
General Course Level	22.53	23	5.68
STEM Course Level	19.05	19	5.35
Mathematics Course Level	14.1	14.5	3.85

Therefore, lower means and medians on the LBHEM Scale Composite Score correspond to more desirable outcomes in general. The descriptive statistics provided in the table indicate that the mathematics course level subgroup had the most desirable outcomes on the scale (mean of 14.1, median of 14.5), followed respectively by the STEM level (mean of 19.05, median of 19), general level (mean 22.53, median of 23), and finally by the developmental level (mean of

23, median of 25). To determine whether at least one of the mathematics course level subgroups differed significantly from one another with respect to their LBHEM Scale Composite Score, a Kruskal-Wallis test was conducted. A significant difference was found between the LBHEM Scale Composite Scores for participants from different mathematics course level subgroups [$H(3) = 30.320,\ p < .001$] (note that the H-test statistic has been adjusted for ties). Participants from the mathematics course level subgroup had a significantly lower LBHEM Scale Composite Score (indicating that the participants in this subgroup had the most desirable responses overall) than each of the other course level subgroups.

The participants of the mathematics course level subgroup had a significantly lower composite score than those in the STEM group [mean rank difference: -30.320, $z = -2.989, p = .017$], those in the general group [mean rank difference: -50.975, $z = -5.191, p < .001$], and those in the developmental group [mean rank difference: -52.548, $z = -3.722, p = .001$]. Additionally, participants from the STEM subgroup had a significantly lower composite score than those in the general group [mean rank difference: -20.655, $z = -2.715, p = .040$]. These analyses align with the inferential analyses conducted on individual SMPI items in the previous section of this chapter and with the descriptive analyses above. Together, they indicate that individuals in the mathematics course level subgroup reported the least levels of limiting characteristics related to mathematics, and in turn, the most desirable responses overall in comparison to participants from the other course level subgroups. Additionally, the results indicate that the more advanced the level of mathematics courses in which participants are enrolled (i.e., in the course level ordering of developmental, general, STEM, and mathematics) the lesser the levels of limiting characteristics and accordingly the more desirable characteristics are reported.

Academic Habits of Mind (AHM) Scale In the analysis of individual SMPI items from the previous section, no significant differences were found on any items from the AHM scale across the mathematics course level subgroups. Figure 4.2.14 depicts the percentage agreement for each of the items of the AHM Scale across each of the mathematics course level subgroups. The data table used to make this graph (which includes the percentage disagreement to each item) is given in Appendix C (Table 4.2.14).

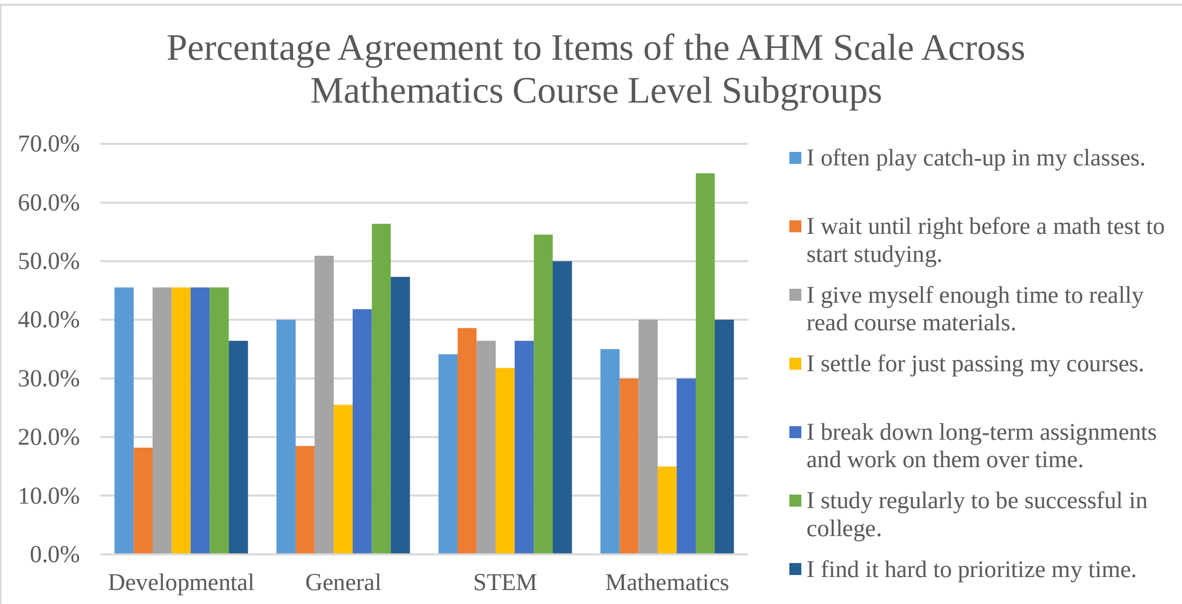

Figure 4.2.14: Graph (Sid-By-Side Bar Graph) Depicting Percentage Agreement to AHM Scale Items Across Mathematics Course Level Subgroups

Although inferential analyses revealed that no significant differences were observed between the items of the AHM scale across the mathematics level subgroups, several descriptive differences and commonalities across the groups should be highlighted. For the item "I settle for just passing my courses" the participants in the mathematics course level subgroup had the most desirable responses (i.e., highest levels of disagreement) with 60% disagreeing, followed by the general level with 54.5%, then by the STEM level with 50%, and finally by the developmental level with 36.4%. However, for the item "I wait until right before a math test to start studying", the students in the mathematics subgroup had the second-highest level of agreement (indicating less desirable outcomes in this regard) with 30%.

The group with the highest level of agreement were the STEM students at 38.6%, followed by the general (18.5%) and developmental (18.2%) levels. Another difference to note is that the developmental level students had the most desirable responses for the item “I break down long-term assignments and work on them over time” agreeing at a rate of 45.5%, followed closely by the general group at 41.8%, then by the STEM group at 36.4%, and the mathematics group at 30%. Therefore, the participants in the mathematics course level subgroup had the least desirable characteristics in this regard. The general group had the highest levels of agreement to the item “I give myself enough time to really read course materials” with 50.9%, followed by the developmental group with 45.5%, then by the mathematics group with 40%, and finally by the STEM group with 36.4%. Taken together, these results indicate that participants in the lower level mathematics course level subgroups have more desirable results with respect to academic habits in several areas, and that in some, the participants in the mathematics course level subgroup have the least.

In terms of commonalities, over a third of the participants in each group agreed to the item “I often play catch-up in my classes” (36.4% of the developmental group, 40% of the general group, 34.1% in the STEM group, and 35% in the mathematics group). A large proportion of participants in each group also agreed to the item “I find it hard to prioritize my time” (36.4% of the developmental group, 47.3% in the general group, 50% of the STEM group, and 40% of the mathematics group). These findings indicate that participants in each of the groups report less desirable academic habits related to time management and study skills. However, a large proportion of the participants in each group agreed to the item “I study regularly to be successful in college" (65% of participants in the mathematics group, 56.4% of the general group, 54.5% of the STEM group, and 45.5% of the developmental group).

To further investigate the AHM scale, descriptive statistics for the AHM Scale Composite Score are given below in Table 4.2.15. Recall from above that the composite score was obtained by summing the responses to each item of the AHM Scale. Since the AHM Scale measures desirable characteristics related to academic habits of mind, higher means and medians on the AHM Scale Composite Score correspond to more desirable outcomes in general.

Table 4.2.15

Descriptive Statistics of AHM Scale Composite Score Across Entire Sample and Mathematics Course Level Subgroups

Subgroup	AHM Scale Composite Score		
	Mean	Median	Standard Deviation
All Participants	22.09	22	4.93
Developmental Course Level	21.36	21	4.57
General Course Level	22.72	22.5	4.52
STEM Course Level	21.47	22.5	4.92
Mathematics Course Level	22.15	23	6.21

The descriptive statistics of the AHM Scale Composite Score indicate that the participants in the general level (mean of 22.72, median of 22.5) reported the most desirable academic habits of mind with respect to the mean (the median of the general and STEM group were identical), followed by the mathematics course level (mean 22.15, median of 23), then by the STEM course level (mean of 21.47, median of 22.5), and finally by the developmental course level (mean of 21.36, median of 21).

Only the mathematics and general level participants had mean and median composite scores higher than the average for all participants. However, it should be noted hat the highest degree of variability was observed in the mathematics group (standard deviation of 6.21). In general, each of the mean and median composite scores for each group were relatively close, indicating that the participants in each group did not differ greatly with respect to the quality of their academic habits. This finding was confirmed inferentially as well. To determine whether at least one of the mathematics course level subgroups differed significantly from one another with respect to their AHM Scale Composite Score, a Kruskal-Wallis test was conducted. No significant differences were found between the AHM Scale Composite Scores for participants from different mathematics course level subgroups $[H(3) = 3.995,\ p = .756]$ (note that the H-test statistic has been adjusted for ties).

Future Orientation (FO) Scale In the analysis of individual SMPI items from the previous section, no significant differences were found on any items from the FO scale across the mathematics course level subgroups. Figure 4.2.16 depicts the percentage agreement to each item of the FO Scale across each of the mathematics course level subgroups. The data table used to make this graph (which includes the percentage disagreement to each item) is given in Appendix C (Table 4.2.16).

One potential limitation of the FO scale that was previously noted in a previous section is that the vast majority of all participants who took the SMPI agreed to each of the items of the FO scale (this is discussed in Chapter 5). This phenomenon was also observed when viewing the results of these items across each of the mathematics course level subgroups. However, it should be noted that for the items “I am taking the necessary steps to obtain the career I desire” and “Thinking bout the future I want makes me do more now to get that future” the mathematics course level subgroup had the lowest levels of agreement. On the item “I have a fairly clear idea of what I need to study now in order to have the career I want”, with the exception of the developmental group (which had 45.5% agreement), the mathematics course level subgroup again had the lowest level of agreement.

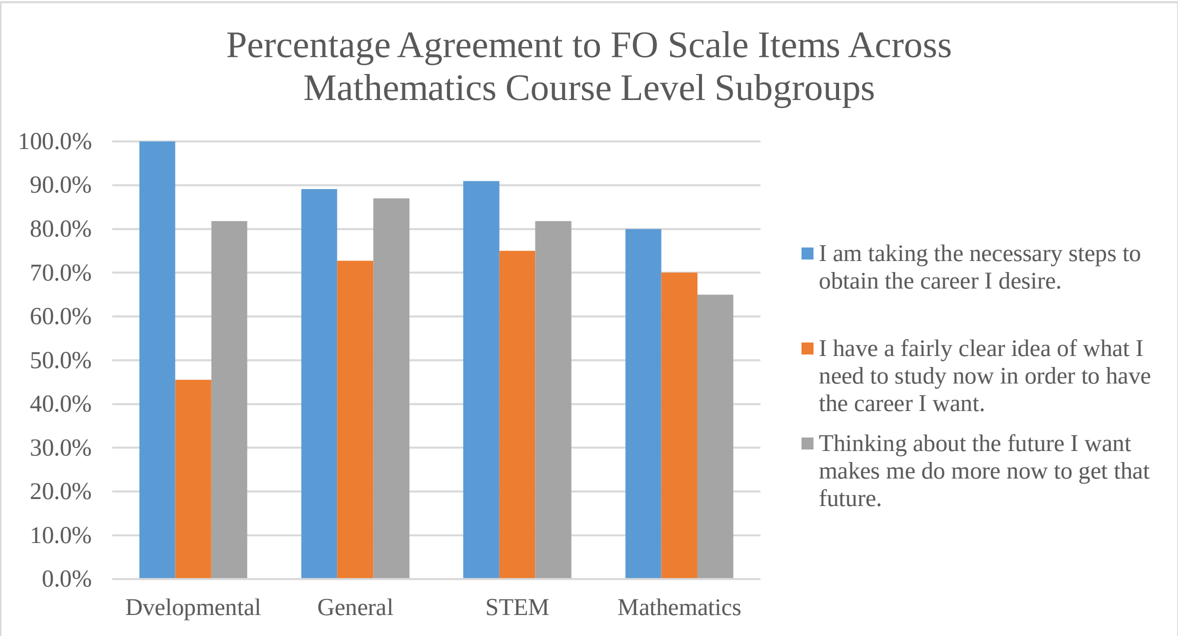

Figure 4.2.16: Graph (Sid-By-Side Bar Graph) Depicting Percentage Agreement to FO Scale Items Across Mathematics Course Level Subgroups

To further investigate the FO scale, descriptive statistics for the FO Scale Composite Score are given below in Table 4.2.17. Recall from above that the composite score was obtained by summing the responses to each item of the FO Scale. Since the FO Scale measures desirable characteristics related to future orientation, higher means and medians on the composite score indicate more desirable outcomes. Both the developmental and mathematics course level subgroups had mean composite scores lower than that of the mean for the entire sample.

Table 4.2.17

Descriptive Statistics of FO Scale Composite Score Across Entire Sample and Mathematics Course Level Subgroup

Subgroup	FO Scale Composite Score		
	Mean	Median	Standard Deviation
All Participants	11.99	12	1.84
Developmental Course Level	10.81	11	2.52
General Course Level	12.33	12	1.72
STEM Course Level	12.09	12	1.44
Mathematics Course Level	11.5	12	2.26

Although each of the descriptive statistics for the FO Scale Composite Score were relatively similar across all groups, the general level group had the highest score (mean of 12.33, median of 12), followed by the STEM course level (mean of 12.09, median of 12), the mathematics course level (mean of 11.5, median of 12), and finally by the developmental course level (mean of 10.81, median of 11).

To determine whether at least one of the mathematics course level subgroups differed significantly from one another with respect to their FO Scale Composite Score, a Kruskal-Wallis test was conducted. No significant differences were found between the AHM Scale Composite Scores for participants from different mathematics course level subgroups [$H(3) = 5.604,\ p = .133$] (note that the H-test statistic has been adjusted for ties).

Academic Self-Concept (ASC) Scale In the analysis of individual SMPI items from the previous section, no significant differences were found on any items from the ASC scale across the mathematics course level subgroups. Figure 4.2.18 below depicts the percentage agreement to each of the ASC Scale items across each of the mathematics course level subgroups. The data table used to make this graph (which includes the percentage disagreement to each item) is given in Appendix C (Table 4.2.18).

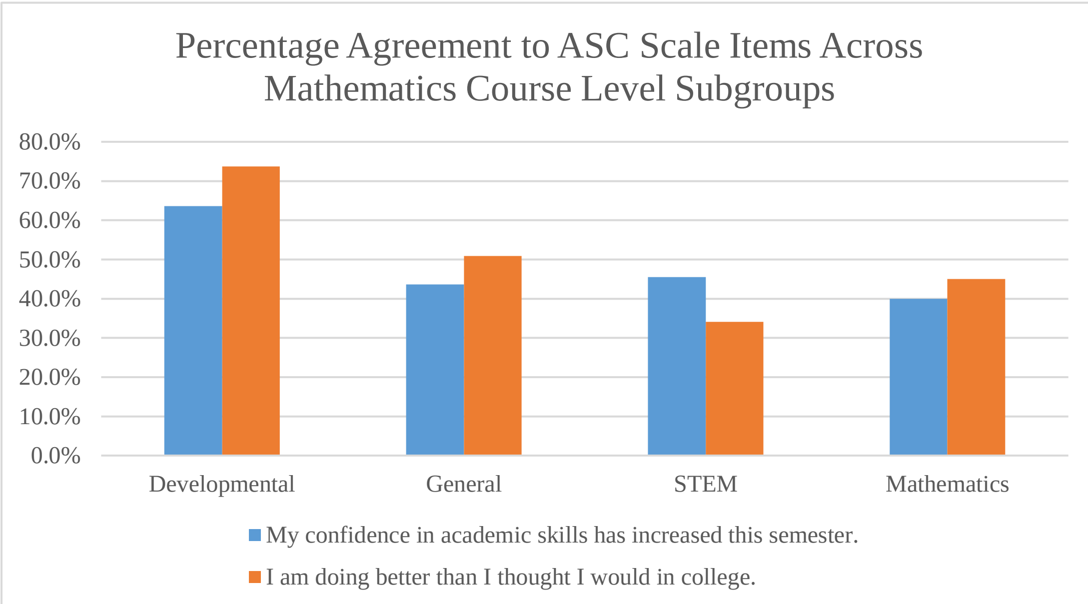

Figure 4.2.18: Percentage Agreement to ASC Scale Items Across Mathematics Course Level Subgroups

A large percentage of each group (more than a third) agreed to both items of the ASC scale. For the first item, “My confidence in academic skills has increased this semester”, the participants from the developmental group had the highest levels of agreement with 63.6%, followed by the general group with 43.6%, then by the STEM group with 45.5%, and lastly by the mathematics group. For the item “I am doing better than I thought I would in college”, the highest level of agreement was again from the developmental group with 72.7%, followed again by the general group with 50.9%, then by the mathematics group with 43.1%, and lastly by the STEM group with 34.1%. To further investigate the ASC scale, descriptive statistics for the ASC Scale Composite Score are given below in Table 4.2.19. Recall from above that the composite score was obtained by summing the responses to each item of the ASC Scale. Since the ASC

Scale measures desirable characteristics related to academic self-concept, higher means and medians on the composite score indicate more desirable outcomes.

Table 4.2.19

Descriptive Statistics of ASC Scale Composite Score Across Entire Sample and Mathematics Course Level Subgroup

Subgroup	ASC Scale Composite Score		
	Mean	Median	Standard Deviation
All Participants	6.67	7	1.84
Developmental Course Level	7.45	8	1.43
General Course Level	6.85	7	1.79
STEM Course Level	6.39	6	1.96
Mathematics Course Level	6.35	6	1.87

The descriptive statistics in Table 4.2.19 indicate that the most desirable responses related to academic self-concept were reported by successively lower subgroups on the mathematics course level subgroup structure. That is, the developmental subgroup had the highest mean and median ASC Scale Composite Score (mean of 7.45, median of 8), followed by the general course level (mean of 7.45, median of 7), then by the STEM course level (mean of 6.39, median of 6), and lastly by the mathematics course level (mean of 6.35, median of 6). One possible interpretation of this is that participants at the lower levels of the course structure may have had lower expectations for themselves and their work in college, and are improving and performing at levels that they did not expect.

To determine whether at least one of the mathematics course level subgroups differed significantly from one another with respect to their ASC Scale Composite Score, a Kruskal-Wallis test was conducted. No significant differences were found between the ASC Scale Composite Scores for participants from different mathematics course level subgroups [$H(3) = 3.995,\ p = .262$] (note that the H-test statistic has been adjusted for ties).

Hesitancy to Engage with Faculty (HEF) Scale In the analysis of individual SMPI items from the previous section, significant differences were found on each of the items of the HEF across the mathematics course level subgroups. Figure 4.2.20 below depicts the percentage agreement to each of the items of the HEF Scale across each of the mathematics course level subgroups. The data table used to make this graph (which includes the percentage disagreement to each item) is given in Appendix C (Table 4.2.20).

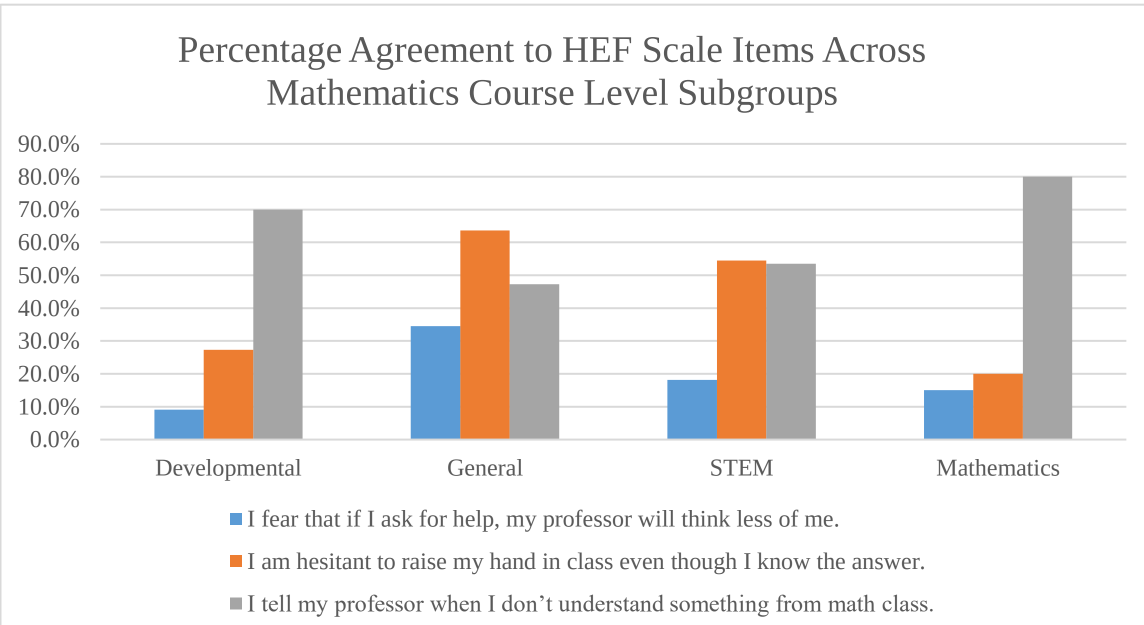

Figure 4.2.20: Percentage Agreement to HEF Scale Items Across Mathematics Course Level Subgroups

For each item, the developmental and mathematics course level subgroups had the most desirable responses. Disagreement to the item "I fear that if I ask for help, my professor will think less of me" indicates a desirable response, and the developmental group had the highest percentage of disagreement to this item with 81.8%, followed closely by the mathematics group with 80%, then by the STEM level group with 65.9%, and finally by the general level group 47.3%. The general group also saw the highest level of agreement with this item (indicating less desirable responses) with 34.5%.

For the item “I am hesitant to raise my hand in class even though I know the answer”, higher levels of disagreement also indicate a desirable response. On this item, the mathematics course level subgroup had the highest percentage of disagreement with 60%, followed by the developmental group (54.5%), then by the STEM group (27.3%), and then by the general level group (23.6%). It should be noted that in addition to having the lowest levels of disagreement to the item, the general and STEM level groups also had the largest levels of agreement to the item. A sizeable majority of participants in the general level group (63.6%) agreed to the item, and a majority of the participants in the STEM level group did as well (54.5%). For the last item of the scale, “I tell my professor when I don’t understand something from math class”, higher levels of agreement are more desirable. Again, the subgroup with the highest level of agreement was the mathematics course level subgroup with 80%, followed again by the developmental subgroup with 70%, then by the STEM level group with 53.5%, and finally by the general level group with 47.3%.

To further investigate the HEF scale, descriptive statistics for the HEF Scale Composite Score are given in Table 4.2.21 below. Recall from above that the composite score was obtained by summing the responses to each item of the HEF Scale. Since the HEF Scale measures undesirable characteristics related to the hesitancy to engage with faculty, lower means and medians on the composite score indicate more desirable outcomes.

Table 4.2.21

Descriptive Statistics of HEF Scale Composite Score Across Entire Sample and Mathematics Course Level Subgroup

Subgroup	HEF Scale Composite Score		
	Mean	Median	Standard Deviation
All Participants	8.07	8	3.15
Developmental Course Level	5.9	5	2.88
General Course Level	8.98	9	3.03
STEM Course Level	8.40	8	2.80
Mathematics Course Level	5.95	5	3.02

The descriptive statistics in Table 4.2.21 indicate that the developmental course level subgroup had the most desirable responses related to hesitancy to engage in faculty (i.e., the lowest level of agreement overall), followed closely by the mathematics course level subgroup. The developmental course level subgroup had the lowest mean score (mean of 5.9, median of 5), followed by the mathematics course level (mean of 5.95, median of 5), followed by the STEM level (mean of 8.4, median of 8), and finally by the general course level (mean of 8.98, median of 9). These results suggest that the developmental and mathematics course level subgroups differ from the other two subgroups with respect to their hesitancy to engage with faculty. This finding was confirmed inferentially.

To determine whether at least one of the mathematics course level subgroups differed significantly from one another with respect to their HEF Scale Composite Score, a Kruskal-Wallis test was conducted. A significant difference was found between the HEF Scale Composite Scores for participants from different mathematics course level subgroups [$H(3) = 19.031,\ p < .001$] (note that the H-test statistic has been adjusted for ties). Participants from the developmental course level subgroup had a significantly lower HEF Scale Composite Score than those in the general level [mean rank difference: -35.832, $z = -2.822,\ p = .029$]. Participants from the mathematics course level subgroup had a significantly lower HEF Scale Composite Score than those in the general level [mean rank difference: -35.707, $z = -3.702,\ p = .001$] and those in the STEM level [mean rank difference: -29.446, $z = -2.946, p = .019$].

Taken together, the results from the analyses on each of the scales of the SMPI across mathematics course level subgroups indicate that participants differ significantly in terms of their reported limiting characteristics in mathematics and their hesitancy to engage with faculty. For the former, the differences align with the hierarchy of the course leveling structure itself, where students in the highest level of the structure, the mathematics course level subgroup, differ significantly from all other groups on every item of the LBHEM scale and the scale composite score overall. Differences were also found between higher course level subgroups and those at lower levels, i.e., participants in the STEM level had more desirable responses than those in the general and developmental level, and those in the general level had more desirable responses than those in the developmental level.

For the latter, the developmental and mathematics course level subgroups had more desirable responses on the items of the HEF than other course subgroups. The mathematics course level subgroups differed significantly from the STEM and general level subgroups and the developmental level differed significantly from the general level in this regard. However, the mathematics and developmental level groups did not differ significantly from one another.

4.2.7 Analysis: SMPI Scale Analysis Across Other Subgroups

Gender Subgroups: Males and Females To determine whether males and females differed significantly each with respect to the SMPI Scale Composite Scores, a set of Mann-Whitney U-Tests were conducted. A significant difference was found between males and females on the AHM Scale Composite Score [$U = 1297.5,\ z = -2.893,\ p = .004$]. Females (mean rank: 70.87) had significantly higher (and thus, more desirable) AHM Composite Scores than males (mean rank: 51.53). This aligns with the findings from Analysis 2.1.2, which indicated that females had more desirable responses on the SMPI items from the AHM Scale. No other significant differences were found between males and females on any of the other scale composite scores.

Academic Major Subgroups To determine whether at least one of the academic major subgroups differed significantly from one another with respect to the SMPI Scale Composite Scores, a set of Kruskal-Wallis tests were conducted. A significant difference was found between different academic major subgroups on the LBHEM Scale Composite Score [$H(5) = 30.798$, $p < .001$]. Pairwise comparisons between the groups revealed that participants in the mathematics major had significantly lower (and thus more desirable) scores on the LBHEM Scale Composite than: those in the science majors [mean rank difference: $-37.531,\ z = 3.963$, $p = .002$], those in the education majors [mean rank difference: $-44.459,\ z = -3.530,\ p = .006$], and those in the health and human services majors [mean rank difference: $-54.502,\ z = -5.326,\ p < .001$]. No other pairwise comparisons across the academic major subgroups were significant. A significant difference was also found between the academic major subgroups on the FO Scale Composite Score [$H(5) = 11.596,\ p = .041$] (the H-statistic was adjusted for ties). Pairwise comparisons revealed that participants in the education major had significantly higher (thus, more desirable) FO Composite Scores than those in the business majors [mean rank difference: $63.479,\ z = 3.330,\ p = .013$]. No other pairwise comparisons across the academic major subgroups were significant. As for the other Scale Composite Scores, although a significant difference was found between the academic year subgroups on the HEF factor, after the Bonferroni correction was applied to the significance values for multiple tests, no pairwise comparisons across the subgroups were significant. The Kruskal-Wallis test on the AHM Scale Composite Score was also not significant.

Academic Year Subgroups To determine whether at least one of the academic year subgroups differed significantly from one another with respect to the SMPI Scale Composite Scores, a set of Kruskal-Wallis tests were conducted. A significant difference was found between academic year subgroups on the HEF Scale Composite Score [$H(3) = 10.616,\ p = .014$] (H-statistic was adjusted for ties). Pairwise comparisons revealed that participants in the Senior academic year subgroup had significantly lower (and thus, more desirable) HEF Scale Composite Scores than those in the Sophomore subgroup [mean rank difference: $-32.927,\ z = -3.126,\ p = .011$]. No other pairwise significant differences were found. Additionally, no other significant differences were found between any of the academic year subgroups on any of the other SMPI Scale Composite Scores.

Mathematics Courses Taken in Senior Year of High School To determine whether at least one of the senior year high school mathematics course subgroups differed significantly from one another with respect to the SMPI Scale Composite Scores, a set of Kruskal-Wallis tests were conducted. A significant difference was found between the different senior year high school mathematics course subgroups on the LBHEM Scale Composite Score [$H(4) = 23.580,\ p < .001$] (H-statistic was adjusted for ties). Pairwise comparisons revealed that participants who reported taking an advanced level mathematics course their senior year had lower LBHEM Scale Composite Scores (thus, more desirable responses overall) than: those who reported taking algebra courses [mean rank difference: $-43.250,\ z = -4.388,\ p < .001$], those who reported taking general courses [mean rank difference: $-27.810,\ z = -3.007,\ p = .026$], and those who reported taking precalculus courses [mean rank difference: $-31.656,\ z = -3.648,\ p = .003$]

their senior year of high school. No other pairwise comparisons were significant. Additionally, no other significant differences were found between any of the senior year mathematics course subgroups on any of the other SMPI Scale Composite Scores.

4.3 Research Question 3: What is the relationship between the performance of students from different collegiate mathematics courses on the modified version of the Cognitive Reflection Test and their responses to a psychoeducational survey measuring domain-general academic habits of mind, future orientation, and academic self-concept, and mathematics-specific academic habits of mind, beliefs, and past academic experiences?

4.3.1 Analysis: Correlational Analyses of MCRT and SMPI Items

To analyze the association between participants' responses to individual items of the SMPI and their scores on the MCRT and CRT subset of the MCRT, Spearman Correlation Coefficient (rho) analyses were run. Recall that due to the ranked nature of the SMPI item data and the MCRT data, the Spearman correlation was chosen as the appropriate analysis to determine the association between these variables. Recall also that like the Pearson correlation, the Spearman correlation measures the strength and direction of the association between two variables. The measure is standardized and can be interpreted in the same fashion as the Pearson correlation, since the coefficient takes on values between -1 and 1. Since the following section will detail correlational analyses on individual subgroups of differing sizes, it should be noted that the significance value assigned to a particular Spearman correlation coefficient depends on the size of the group under analysis (for a full description of the calculation and assignment of significance values, see IBM, 2013, p. 641). That is, depending on the size of the group under analysis, a correlation coefficient that is significant in one analysis may not be significant for another analysis if the groups differ in size.

4.3.2 Analysis: Correlational Analysis of MCRT Scores and SMPI Responses Across Entire Sample

A Spearman correlational analysis was conducted between each of the 33 Likert scale items of the SMPI that measure psychoeducational facets and both the MCRT scores and CRT subset scores of the MCRT for all 130 participants. The statistically significant correlations are given below in Table 4.3.1. In the table, two columns represent the MCRT and CRT data, and each of the rows represent the data from the individual items of the SMPI. Each row and column intersects at a cell, and the correlation between the MCRT or CRT data (the columns) and the individual SMPI item data is given in that cell. For example, given in the cell that is in the MCRT column and the "I'm just not good at math" row is the Spearman correlation coefficient (rho) that was calculated between the two sets of data. For the purposes of making the table readable and succinct, only the statistically significant Spearman correlations are given. Correlation coefficients in the table indicated with a superscript (*) indicate significance at the $\alpha = .05$ level, and coefficients in the table indicated with a double superscript (**) indicate significance at the $\alpha = .01$ level. If a correlation is significant for one of the two MCRT scores and not for the other, than only the significant correlation is given and the other is omitted.

The strongest correlations that were observed between the SMPI items and the MCRT and CRT scores were those related to mathematics. In particular, the correlations between all 7 items of the LBHEM scale and both the MCRT and CRT scores were statistically significant at the $\alpha = .01$ level. The strongest correlation observed was that between the MCRT and CRT scores and the item "I can explain how I reach the correct answer on a math test".

Table 4.3.1

Correlations Between MCRT and CRT Scores and Individual SMPI Items Across the Entire Sample

SMPI Item	MCRT Score	CRT Score
I'm just not good at math.	-.452**	-.418**
The wording of math problems confuses me.	-.384**	-.393**
I use my time between classes productively	-.177*	-
I settle for just passing my courses.	-.197*	-
I break down long-term assignments and/or class projects and work on them over time.	-.229*	-.181*
Math and/or anything with numbers has been an obstacle to my academic success.	-.420**	-.498**
I expect to use the math I have learned in my future career.	.354**	.342**
I have experienced difficulties in math since high school or before high school.	-.374**	-.385**
I am always well-prepared for math class.	.189*	.230*
I am usually confident that I will do well on math tests.	.313**	.276**
I push aside math assignments and do them last.	-.300**	-.256**
I can explain how I reach the correct answer on a math test.	.500**	.455**
Sometimes, I am disappointed in my test results because I studied a great deal.	-.334**	-.324*
My grades in college math have influenced what degree I can pursue.	-.188*	-.177*
When something confuses me, I think about it until I can make sense out of it.	-	.206*
I am always well prepared for math class.	.189*	.230**

This indicates that those who can conceptualize about solution methods and communicate them effectively tend to perform better on the MCRT and CRT. The next strongest items were "I'm just not good at math" and "Math and/or anything with numbers has been an obstacle to my academic success", with the latter having the strongest correlation overall with the CRT. The correlations, respectively, between both of these items and the MCRT and CRT scores were negative. This suggests that those with more self-limiting beliefs and experiences related to mathematics tended to perform lower on the MCRT and CRT; conversely, this indicates that those with less self-limiting beliefs and experiences related to mathematics tended to perform better on the MCRT and CRT. This was also true of the relationship between the item "I have experienced difficulties in mathematics since high school or before high school" and the MCRT and CRT scores. Another item for which the correlation was negative, was "The wording of math problems confuses me," which indicates that those who are confused by the wording of math problems tended to have lower scores on the MCRT and CRT. This is consistent with the nature of the problems of the MCRT, which are each word problems. Moreover, the wording of the CRT problems involved a language component that required inhibitory control to overcome. It should be noted that the correlation between this SMPI item and the CRT score was stronger than that of the item and the MCRT.

Other significant correlations related to mathematics include "I am usually confident that I will do well on math tests." The correlation here was positive indicating that those who have greater confidence on math tests tended to exhibit better performance on the MCRT and CRT. Another LBHEM item, "I push aside math assignments and do them last" was also negatively associated with MCRT and CRT scores, indicating those who actively avoid mathematics assignments tended to score lower on the MCRT and CRT. Other items that were related to mathematics, but were not on the LBHEM Scale, were also significantly associated with the MCRT and CRT scores. The item "I expect to use the math I have learned in my future career" was positively associated with the MCRT and CRT scores, indicating that an expectation to use mathematics in the future was positively related to success on the MCRT and CRT. The item "My grades in college math have influenced what degree I can pursue" was negatively associated with the MCRT and CRT scores, indicating that negative experiences in college mathematics and their implications were associated with lower scores on the assessments. Lastly, the item "I am always well-prepared for math class" was also significantly correlated with MCRT and CRT scores, and the relationship was stronger with the CRT than the MCRT.

Several items related to general academic habits and experiences were also significantly related to the MCRT and CRT scores. The strongest of these was "Sometimes, I am disappointed in my test results because I studied a great deal", which was negatively related to the MCRT and CRT scores, indicating that those who experience disappointment despite committing themselves to studying tended to score lower on the MCRT and CRT. Additionally, the responses to the item "I break down long-term assignments and/or class projects and work on them over time" was negatively associated with MCRT and CRT scores. This would suggest that those who are able to manage their time effectively and manage long-term assignments tend to have lower scores on the MCRT and CRT. The nature of this finding is unclear, and will be the subject of further research on the SMPI (see Chapter 5 for a discussion of this).

Two items related to general time management and academic habits of mind were significantly related to the MCRT but not the CRT. These were "I often play catch-up in my classes" and "I use my time between classes productively," and both were negatively correlated with the MCRT and CRT scores. For the former, this suggests that those who are often behind in their classes academically tend to perform at a lower level on the MCRT and CRT scores. For the latter, this finding implies that those who report using their time effectively between classes tend have lower scores on the MCRT and CRT. Much like the finding above related to the item "I break down long-term assignments and/or class projects and work on them over time", the nature of this relationship is also unclear and will be the subject of further research on the SMPI (see Chapter 5 for further discussion of this). Finally, one item related to academic habits that was related to the CRT but not the MCRT was "When something confuses me, I think about it until I can make sense out of it." This implies that those who think reflectively and persistently about difficult material tend to have better scores on the CRT. This is consistent with the nature of the CRT problems, which require Type 2 thinking (which is more reflective and effortful) in order to be successful.

4.3.3 Analysis: Correlational Analysis of MCRT Scores and SMPI Responses Across Mathematics Course Level Subgroups

In the previous analysis involving all participants, several items were identified as being significantly correlated with MCRT and CRT scores. These items were further investigated across each of the mathematics course level subgroups. Correlational analyses on all the SMPI items and MCRT and CRT scores were conducted on each mathematics course level subgroup. That is, for each of the following analyses, the responses for the participants in each subgroup were analyzed separately to determine differences in the correlations across each subgroup. Recall that the developmental level had 11 participants, the general level had 55 participants, the STEM level had 44 participants, and the mathematics level had 20 participants. Table 4.3.2 below includes the respective Spearman correlation coefficients calculated for each subgroup's LBHEM Scale item responses and the MCRT and CRT scores. In the table, the SMPI item that is being analyzed in comparison to MCRT and CRT scores is given in the first column. The second column indicates each of the mathematics course level subgroups, and the final two columns indicate the MCRT and CRT scores.

Therefore, in the cells of the table at intersections of each row (which indicate participants' SMPI responses from each subgroup) and the MCRT and CRT columns, the Spearman correlation coefficients are given. Correlation coefficients in the table indicated with a superscript * indicate significance at the $\alpha = .05$ level, and coefficients in the table indicated with a double superscript ** indicate significance at the $\alpha = .01$ level. If a correlation is significant for one of the two MCRT scores and not for the other, than only the significant correlation is given and the other is omitted.

For example, the first SMPI item is “I’m just not good at math”, and the responses to this item are separated across the different mathematics course level subgroups, and the correlations between these responses and the MCRT and CRT scores are given. The intersection of the row indicated by “Developmental Level” and the column indicated “MCRT Score” contains the value -.621*. This communicates that the correlation between the responses to “I’m just not good at math” from the participants in the developmental level subgroup and their MCRT scores is -.621, which is significant at the $\alpha = .05$ level. Recall that the significance value assigned to a particular Spearman correlation coefficient depends on the size of the group under analysis. That is, depending on the size of the group under analysis, a correlation coefficient that is significant in one analysis may not be significant for another analysis if the groups differ in size.

For the item “I’m just not good at math”, the correlations for each subgroup were negative, and the strongest of these for the MCRT was observed for the developmental group. However, for the CRT, the strongest correlation was observed by the STEM group, followed by the general group, both of which were significant. This suggests that this self-limiting belief is more strongly related to the CRT performance of the STEM and general level participants than those in the other subgroups. The correlations were the smallest for the mathematics level (note that the correlations for this group for the MCRT and CRT are the same since the scores for both assessments for the participants in this group are perfectly correlated), and much smaller in comparison to the other groups, indicating that such a limiting belief is not strongly related to their outcomes on the MCRT and CRT.

Table 4.3.2

Correlations Between MCRT and CRT Scores and LBHEM Scale Items Across Mathematics Course Level Subgroups

SMPI Item	Subgroup	MCRT Score	CRT Score
I'm just not good at math.	Developmental Level	-.621*	-.318
	General Level	-.288*	-.328*
	STEM Level	-.404**	-.375*
	Mathematics Level	-.064	-.064
The wording of math problems confuses me.	Developmental Level	-.379	-.319
	General Level	-.035	.115
	STEM Level	-.423**	-.527**
	Mathematics Level	-.226	-.226
Math and/or anything with numbers has been an obstacle to my academic success.	Developmental Level	-.513	-.530
	General Level	-.307*	-.431*
	STEM Level	-.374*	-.521**
	Mathematics Level	.081	.081
I have experienced difficulties in math since high school or before high school.	Developmental Level	-.505	-.313
	General Level	-.108	-.301*
	STEM Level	-.222	-.159
	Mathematics Level	-.229	-.229
I am usually confident that I will do well on math tests.	Developmental Level	.423	.415
	General Level	.238	.375**
	STEM Level	.348*	.326*
	Mathematics Level	.378	.378
I can explain how I reach the correct answer on a math test.	Developmental Level	.529	.392
	General Level	-.313*	.328*
	STEM Level	.544**	.477**
	Mathematics Level	.011	.011
I push aside math assignments and do them last.	Developmental Level	-.740**	-.323
	General Level	-.122	-.229
	STEM Level	-.291	-.204
	Mathematics Level	.251	.251

However, it is also the case that the mathematics course level subgroup scored the highest on both the MCRT and CRT and had the least levels of agreement to this item. This pattern was also observed for the items "Math and/or anything with numbers has been an obstacle to my academic success" where the mathematics level subgroup had the lowest correlations in comparison, and the STEM and general level groups had significant negative correlations with both MCRT and CRT performance. The strength of these correlations was greater for the CRT than the MCRT. The correlation was also similarly-sized for the developmental subgroup, but was not significant (this is likely due to the small sample size of this group).

The responses to the item, “I can explain how I reach the correct answer on a math test” also followed a similar pattern, where the STEM and general level participants’ responses were significantly correlated with their MCRT and CRT scores; however in this instance the correlations were positive. The developmental group’s responses were also correlated at a similar strength but were again not significant. Additionally, the mathematics course level subgroup’s participants again had the weakest correlations overall. The findings from these correlations suggest that self-limiting beliefs, past experiences, and ability to explain mathematical work are more strongly related to success on the MCRT and CRT for the developmental, general, and STEM level participants than for those in the mathematics course level subgroup. This could also be due to the fact that those in the mathematics subgroup reported the lowest levels of these characteristics in comparison to the other groups.

The correlations between the responses to the item “The wording of math problems confuses me” were significant for the STEM level group and the strongest overall with both the MCRT and CRT. The correlations for all but the general level group were similarly sized and negative, indicating that the difficulty associated with the wording of math problems was associated with lower performance on the MCRT and CRT. For the general group, their responses to the item were negatively correlated with the MCRT and positively correlated with the CRT. This would indicate that having a difficulty deciphering word problems would be associated with better performance on the CRT, which is inconsistent and likely due to the fact that the magnitude of the correlations for this group were so low. The correlations indicate that there is a relationship between this difficulty and performance on the MCRT and CRT for participants in the mathematics course level subgroup as well. .

For the item, “I have experienced difficulties in math since high school or before high school” all the correlations were negative and only the general level group’s responses were significantly correlated with the CRT. The correlations for the other groups were similar in size, with the STEM level having the lowest for the CRT and the general level having the lowest for the MCRT. The strength of these correlations were lower than those (with the only exception being the mathematics level subgroup) for the item “Math and/or anything with numbers has been an obstacle to my academic success”. This may indicate that participants view having experienced long-term difficulties in mathematics and having mathematics be an obstacle to success differently. One possible interpretation of this is that participants may be viewing difficulties as being related to challenging content, and not necessarily forming an obstacle to success.

For the item, “I push aside math assignments and do them last” the correlations for each of the groups were negative, with the exception of the mathematics course level subgroup. The positive correlations observed for the mathematics course level subgroup indicate that actively avoiding mathematics work is positively related to success on the MCRT and CRT. However, this finding is likely due to the fact that participants in the mathematics course level subgroup had the lowest levels of agreement to this item and the best scores on both the MCRT and CRT. For each of the other groups, the correlations were similarly-sized in relation to the CRT. The strongest correlation for this item was observed between the responses of the developmental level group and the MCRT. This could be due to the fact that 90.9% of developmental level participants had a zero score on the CRT, and only earned points on the MCRT. For the item "I am usually confident that I will do well on math tests", all of the correlations were positive and were similarly-sized for each group. The correlation between the responses of the STEM group were significantly correlated with both the MCRT and the CRT at the $\alpha = .05$ level, and the correlation between the responses for the general level group to this item were significantly correlated with their CRT scores at the $\alpha = .01$ level. This finding suggests that the confidence with which participants approach mathematics assessments was positively related to success on the MCRT and CRT and this relationship was relatively consistent across each subgroup.

Above, each of the items of the LBHEM were investigated across ach of the mathematics course level subgroups since they emerged as significantly correlated to MCRT and CRT success for all participants. Other items not on the LBHEM were also significantly correlated with MCRT and CRT scores—these are detailed in Table 4.3.3 below. The structure of Table 4.3.3 is the same as for Table 4.3.2 above, where the correlations are given between the indicated SMPI item and MCRT and CRT scores for all mathematics course level subgroups.

Although they are not on the LBHEM Scale, three items in the table are related to mathematics. For the item "I expect to use the math I have learned in my future career", STEM level participants' responses were significantly positively correlated to both the MCRT and CRT. This suggests that the expectation to use the math they have learned in the future was associated with higher scores on the MCRT and CRT. However, for the other subgroups, such as the general and mathematics subgroups, the responses were either very small in magnitude or negatively correlated. It is unclear why this would be the case, and will be the subject of future SMPI research (see Chapter 5 for a discussion of this). For the item, "My grades in college math have influenced what degree I can pursue", a similar phenomenon was observed where some of the correlations were negative and some were positive.

Table 4.3.3

Correlations Between MCRT and CRT Scores and SMPI Items Not from the LBHEM Scale Across Mathematics Course Level Subgroups

SMPI Item		MCRT Score	CRT Score
I use my time between classes productively	Developmental Level	.321	.524
	General Level	.015	.063
	STEM Level	-.361*	-.319*
	Mathematics Level	-.160	-.160
I settle for just passing my courses.	Developmental Level	-.039	-.052
	General Level	-.222	-.150
	STEM Level	-.269	-.173
	Mathematics Level	.140	.140
I break down long-term assignments and/or class projects and work on them over time.	Developmental Level	-.460	-.258
	General Level	-.139	-.121
	STEM Level	-.156	.064
	Mathematics Level	-.364	-.364
I expect to use the math I have learned in my future career.	Developmental Level	.147	-.053
	General Level	-.082	-.241
	STEM Level	.364*	.447*
	Mathematics Level	-.219	-.219
Sometimes, I am disappointed in my test results because I studied a great deal.	Developmental Level	-.623*	-.531
	General Level	-.309*	-.217
	STEM Level	-.232	-.211
	Mathematics Level	-.370	-.370
My grades in college math have influenced what degree I can pursue.	Developmental Level	.506	.441
	General Level	-.243	-.375**
	STEM Level	-.376*	-.290
	Mathematics Level	.251	.251
I am always well prepared for math class.	Developmental Level	.492	.365
	General Level	.186	.381**
	STEM Level	.044	.124
	Mathematics Level	.057	.057

The significant correlations were observed in the STEM group for the MCRT, and the general level for the CRT, suggesting that negative experiences in college-level math are associated with lower performance on the CRT. However, again, the responses for two groups, the developmental and mathematics group, were positively correlated with MCRT and CRT scores. This could be due to the fact that the mathematics participants had such a high score on the assessments and only one developmental participant earned a non-zero score on the CRT (still, this will be explored further in future research). The last mathematics-related item was "I am always well-prepared for math class" with which the MCRT and CRT scores for all groups were positive, suggesting that preparedness for mathematics class is associated with higher scores on the MCRT and CRT. The correlation was significant for the general level group.

One item above for which a significant correlation was observed were the items "I use my time between classes productively". The correlation for the STEM level group was significant and negative, suggesting that the productive use of time between classes is associated with a lower score on the MCRT and CRT. This finding is similar to the one that the correlations for most groups' responses to the item "I break down long-term assignments and/or class projects and work on the mover time" are negative. The nature of why several items related to academic habits of mind are negatively associated with success on the MCRT and CRT is unclear, and will be explored in future research.

For the item "Sometimes, I am disappointed in my tests results because I studied a great deal", significant negative correlations were observed for both the general and STEM level groups in relation to the MCRT, with the developmental group's being particularly strong. The correlations for all subgroups were negative, indicating that the experience of not meeting one's expectations on assessments is associated with lower scores on the MCRT and CRT. For the mathematics course level subgroup, the correlation was greater in magnitude than each of the other groups on the CRT, but it was not significant (likely due to the size of this group). Other significant correlations did arise for items not listed above. The correlation between the responses of the developmental group to the item “I have a fairly clear idea of what I need to study now in order to have the career I want” was significantly positively correlated with their MCRT and CRT scores ($\rho = .360, \alpha < .05$), suggesting that a future time perspective is associated with higher scores on the MCRT. Also, the responses of the STEM level subgroup’s participants to the item “The hardest part about solving word problems is understanding what is being asked” were significantly negatively correlated with their MCRT scores, indicating that a difficulty understanding word problems is associated with lower MCRT scores. This is consistent with the nature of the MCRT as an assessment consisting of five mathematics word problems.

Taken together, these findings indicate that responses to many of the SMPI items are related to the scores observed on the MCRT and CRT. Significant relationships were often found between the responses of the participants form the general and STEM level groups, particularly for those on the LBHEM Scale, which indicates the relationship between these participants reported limiting characteristics in mathematics and the MCRT and CRT are notable. It should also be noted that no correlations for any LBHEM item were significant for the mathematics course level group. This may indicate several phenomena. First, it may indicate that limiting characteristics for students in this subgroup are not strongly related to their success on the MCRT and CRT. However, the participants from this group reported the lowest levels of limiting characteristics on the SMPI and performed at such a high level on the MCRT and CRT, which may explain the lack of an association between the two sets of data. With respect to the other items of the SMPI that were significantly correlated to the MCRT and CRT for all participants, the analysis across the mathematics course level subgroups indicate that only several of these are so for each subgroup. Additionally, a recurring finding regarding that several items related to general academic habits of mind will be investigated in further research. It is currently unclear why higher levels of academic habits of mind related to time management would be negatively associated with success on the MCRT and CRT.

4.3.4 Analysis: Correlational Analysis of MCRT Scores and SMPI Responses Across Other Subgroups

In the previous analyses involving all participants and the different mathematics course level subgroups, several items were identified as being significantly correlated with MCRT and CRT scores. The first set of items consisted of those from the LBHEM Scale, which were explored further for each of the other subgroups under analysis. The previous analysis explored each of the other times not from the LBHEM scale that were correlated with MCRT and CRT scores for all participants further across each of the other subgroups. In this section, however, only items for which the responses for the other subgroups were significantly correlated with MCRT and CRT scores are discussed further.

Gender Subgroups: Males and Females The table below, Table 4.3.4, includes the respective Spearman correlation coefficients calculated for the responses to each of the items of the LBHEM Scale for the participants in the male and female subgroups and their respective MCRT and CRT scores. The table follows the same structure as the previous tables (Tables 4.3.2 and 4.3.3). Recall, that the female subgroup consisted of 78 participants and the male subgroup consisted of 49 participants.

With the exception of one correlation (between the responses of the males on the item "I push aside math assignments and do them last" and their scores on the CRT), every other correlation in the table was significant. Moreover, of these, all but two correlations (which were significant at the $\alpha = .05$ level) were significant at the $\alpha = .01$ level.

Table 4.3.4

Correlations Between MCRT and CRT Scores and SMPI Items Not from the LBHEM Scale Across Mathematics Course Level Subgroups

SMPI Item		MCRT Score	CRT Score
I'm just not good at math.	Males	-.492**	-.398**
	Females	-.426**	-.461**
The wording of math problems confuses me.	Males	-.391**	-.446**
	Females	-.364**	-.312**
Math and/or anything with numbers has been an obstacle to my academic success.	Males	-.410**	-.457**
	Females	-.348*	-.461**
I have experienced difficulties in math since high school or before high school.	Males	-.401**	-.325**
	Females	-.243*	-.373**
I am usually confident that I will do well on math tests.	Males	.544**	.478**
	Females	.277*	.384**
I can explain how I reach the correct answer on a math test.	Males	.476**	.388**
	Females	.458**	.467**
I push aside math assignments and do them last.	Males	-.297*	-.208
	Females	-.266**	-.293**

Many of the correlations were moderate in strength, and each aligned with the desirability of each item. That is, items that were negatively coded (meaning that higher levels of agreement were associated with less desirable outcomes) were negatively correlated with MCRT and CRT scores, and items that were positively coded (meaning that higher levels of agreement were associated with more desirable outcomes) were positively correlated with MCRT and CRT scores. Overall, for both subgroups, higher levels of reported limiting beliefs, habits, and experiences related to mathematics were associated with lower scores on both the MCRT and CRT.

With respect to the other items of the SMPI several were also significantly correlated with MCRT and CRT scores. Female participants' responses to the item "I settle for just passing my courses" was significantly negatively correlated with both their MCRT ($\rho = -.263, p < .05$) and CRT ($\rho = -.229,\ p < .05$) scores. Female participants' responses to the item "I expect to use the math I have learned in my future career" was significantly positively correlated with both their MCRT ($\rho = .443,\ p < .01$) and CRT ($\rho = .479,\ p < .01$) scores. Additionally, female participants' responses to the item "Sometimes I am disappointed in my test results because I studied a great deal" were significantly negatively correlated with their MCRT ($\rho = -.303,\ p < .05$) and CRT ($\rho = -.350,\ p < .05$) scores. Male participants' scores for this item were also significantly negatively related to their MCRT scores ($\rho = -.317, p < .05$). Female participants' responses to the item "My confidence in academic skills has increased this semester" was significantly positively correlated with their scores on the MCRT ($\rho = .293, p < .01$) and the CRT ($\rho = .351,\ p < .01$). Males' responses to the item "I am always well prepared for math class" were significantly positively correlated with their CRT scores ($\rho = .362,\ p < .01$), indicating that consistent preparedness to attend mathematics class is associated with higher scores on the CRT for males.

Female participants' responses to the item "I have a 'game plan' that is effective for tackling math homework" were significantly positively correlated with their CRT scores ($\rho = .244,\ p < .05$). Finally, male participants' responses to the item "I wait until right before a math test to start studying" was significantly positively correlated with both their MCRT ($\rho = .383$, $p < .01$) and CRT ($\rho = .315,\ p < .01$) scores. This would suggest that the propensity to procrastinate studying for mathematics assessments was associated with better scores on the MCRT and CRT. Unlike each of the significant correlations just described, the nature of this finding is unclear, and will be the subject of further SMPI research (see Chapter 5 for a discussion of directions for future research).

Academic Year Subgroups The table below, Table 4.3.5, includes the respective Spearman correlation coefficients calculated for the responses to each of the items of the

LBHEM Scale for the participants in each of the academic year subgroups and their respective MCRT and CRT scores.

Table 4.3.5

Correlations Between MCRT and CRT Scores and SMPI Items Not from the LBHEM Scale Across Academic Year Subgroups

SMPI Item		MCRT Score	CRT Score
I'm just not good at math.	First-Year	-.442**	-.384**
	Sophomore	-.579**	-.391*
	Junior	-.390*	-.420*
	Senior	-.481*	-.488*
The wording of math problems confuses me.	First-Year	-.013	.045
	Sophomore	-.423*	-.587**
	Junior	-.342*	-.403*
	Senior	-.739**	-.648**
Math and/or anything with numbers has been an obstacle to my academic success.	First-Year	-.519**	-.535**
	Sophomore	-.196	-.535**
	Junior	-.389*	-.441**
	Senior	-.537*	-.530*
I have experienced difficulties in math since high school or before high school.	First-Year	-.170	-.215
	Sophomore	-463*	-.515**
	Junior	-.435*	-.492**
	Senior	-.441*	-.359*
I am usually confident that I will do well on math tests.	First-Year	.244	.324*
	Sophomore	.402*	.219*
	Junior	.479**	.528**
	Senior	.737**	.696**
I can explain how I reach the correct answer on a math test.	First-Year	.391**	.263
	Sophomore	.504**	.452*
	Junior	.474**	.414*
	Senior	.691*	.631*
I push aside math assignments and do them last.	First-Year	-.271	-.240
	Sophomore	-.418*	-.453*
	Junior	-.201	-.100
	Senior	-.572*	-.532*

The table follows the same structure as the previous tables (Tables 4.3.3 and Table 4.3.4). Recall, that the sizes of the subgroups are 46 for the first-year, 26 for the sophomore, 34 for the junior, and 22 for the senior.

With the exception of three correlations involving the first-year subgroup, all of the correlations between the responses to the items of the LBHEM by the participants in each subgroup were significantly correlated with their respective CRT scores, with most significant at the $\alpha = .01$ level. With the exception of three correlations (two involving the first-year students and one involving the sophomore students), all of the correlations between the responses to the items of the LBHEM by the participants in each subgroup were significantly correlated with their respective MCRT scores as well, again with most significant at the $\alpha = .01$ level. With the exception of one correlation (between the first-year participants' responses to the item "The wording of math problems confuses me")
every correlation aligned with the desirability of each item. That is, items that were negatively coded (meaning that higher levels of agreement were associated with less desirable outcomes) were negatively correlated with MCRT and CRT scores, and items that were positively coded (meaning that higher levels of agreement were associated with more desirable outcomes) were positively correlated with MCRT and CRT scores. In addition, most of the observed correlations were moderate in strength. Overall, across all subgroups, higher levels of reported limiting beliefs, habits, and experiences related to mathematics were associated with lower scores on both the MCRT and CRT.

The strongest correlations overall in the table above were observed for the senior subgroup. Two of these correlations was observed between the responses of the senior group participants on the item "The wording of math problems confuses me" and both the MCRT ($\rho = -.739, p < .01$) and the CRT ($\rho = -.648, p < .01$). This finding is consistent with the nature of the MCRT as an assessment consisting of word problems, some of which involve language demands that influence the solving of the problem. Other particularly strong correlations for the senior subgroup were observed for the items "I am usually confident that I will do well on math tests" ($\rho = .737$ with MCRT, $p < .01$, and $\rho = .696$ with CRT, $p < .01$) and "I can explain how I reach the correct answer on a math test ($\rho = .691$ with MCRT, $p < .01$, and $\rho = .631$ with CRT, $p < .01$).

With respect to the other items of the SMPI several were also significantly correlated with MCRT and CRT scores. For the item, "I expect to use the math I have learned in my future career" juniors' responses were significantly positively correlated with both the MCRT ($\rho = .437, p < .01$) and the CRT ($\rho = .406, p < .05$). Seniors' responses to this item were also significantly positively correlated with their MCRT scores ($\rho = .499, p < .05$). For the item "My grades in college math have influenced what degree I can pursue" first year participants' responses were significantly negatively correlated with both their MCRT ($\rho = -.314, p < .05$) and CRT scores ($\rho = -.415,\ p < .01$). Senior participants' responses to the item "I am always well-prepared for math class" were significantly positively correlated with their MCRT ($\rho = .435,\ p < .05$) and CRT ($\rho = .424, p < .05$) scores. Junior participants' responses to the item "I am doing better than I thought I would in college" were significantly negatively correlated with their CRT scores ($\rho = -.341, p < .05$). Lastly, senior participants' responses to the item "The hardest part about solving word problems is understanding what is being asked" were significantly negatively correlated with their MCRT scores ($\rho = -.464, p < .05$), which is consistent with the nature of the MCRT.

Academic Major Subgroups Table 4.3.6 includes the respective Spearman correlation coefficients calculated for the responses to each of the items of the LBHEM Scale for the participants in each of the academic major subgroups and their respective MCRT and CRT scores. The table follows the same structure as previous tables (Tables 4.3.4 and 4.3.5). Recall, that the sizes of the subgroups are 6 students with no major, 5 for business-related majors, 15 for education-related majors, 35 for health and human sciences-related (HHS) majors, 47 science-related majors, and 22 for mathematics majors.

It should be first noted that for the subgroup of participants reporting having no major, all participants scored a zero on the CRT. Therefore, due to this, and due to the fact that the subgroup is so small, Spearman correlations could not be calculated between SMPI items and their score on the CRT. As was observed with the participants who were in the mathematics course level subgroup, the responses of the participants in the mathematics major were not significantly correlated with their MCRT and CRT scores for any of the SMPI items. This was also the case for the education majors, despite many of the correlations between their responses to SMPI items and their MCRT and CRT scores were moderate in strength (this is likely due to the small sample size of the subgroup). Although the smallest subgroup was the business-related majors subgroup, the correlations between their responses to LBHEM SMPI items and their MCRT and CRT scores were the strongest significant correlations overall.

In fact, the strongest correlation in the table was a perfect correlation ($\rho = 1.00$) between the responses of the participants in the business-related majors subgroup and both the MCRT and CRT (the scores of the business majors on both instruments were equivalent) on the item “I can explain how I reach the correct answer on a math test”.

Table 4.3.6

Correlations Between MCRT and CRT Scores and SMPI Items Not from the LBHEM Scale Across Academic Major Subgroups

SMPI Item		MCRT Score	CRT Score
I'm just not good at math.	None	-.220	-
	Education	-.379	-.197
	Business	-.913*	-.913*
	HHS	-.395*	-.377*
	Science	-.350*	-.416**
	Mathematics	-.100	-.034
The wording of math problems confuses me.	None	-.426	-
	Education	-.230	.260
	Business	-.889*	-.889*
	HHS	-.064	.043
	Science	-.377**	-.396**
	Mathematics	-.250	-.348
Math and/or anything with numbers has been an obstacle to my academic success.	None	.433	-
	Education	-.502	-.449
	Business	-.866	-.866
	HHS	-.436**	-.404*
	Science	-.416**	-.532**
	Mathematics	.154	.081
I have experienced difficulties in math since high school or before high school.	None	.107	-
	Education	-.304	-.453
	Business	-.889*	-.889*
	HHS	-.277	-.377*
	Science	-.144	-.165
	Mathematics	-.162	-.133
I am usually confident that I will do well on math tests.	None	-.224	-
	Education	.500	.447
	Business	.968**	.968**
	HHS	.245	.420*
	Science	.300*	.339*
	Mathematics	.358	.321
I can explain how I reach the correct answer on a math test.	None	.746	-
	Education	.164	.406
	Business	1.00**	1.00**
	HHS	.440**	.365*
	Science	.549**	.479**
	Mathematics	.029	.043
I push aside math assignments and do them last.	None	-.335	-
	Education	-.173	-.424
	Business	-.913*	-.913*
	HHS	-.258	-.243
	Science	-.140	-.088
	Mathematics	-.002	.076

This was followed by a $\rho = .968$ correlation with the item "I am usually confident I will do well on math tests", a $\rho = -.913$ on the item "I push aside math assignments and do them last", a ρ =-. 889 on both the items "I have experienced difficulties in mathematics since high school or before high school" and "The wording of math problems confuses me."

Statistically significant correlations between other subgroups' responses and their MCRT and CRT scores were also observed. Participants' responses to the item "I'm just not good at math" were significantly negatively correlated to both MCRT scores for the HHS ($\rho = -.395, p < .05$) and science majors ($\rho = -.350, p < .05$), and the CRT scores for these groups (HHS: $\rho = -.377, p < .05$, science: $\rho = -.416, p < .01$). This was also true for the item "Math and/or anything with numbers has been an obstacle to my academic success" where the responses for the HHS students were significantly correlated with their MCRT (ρ =-.377, p<.01) and CRT ($\rho = -.404, p < .05$), as was the case with the science majors (MCRT: $\rho = -.416, p < .01$, CRT: $\rho = -.532, p < .01$). Theis pattern was again observed between the HHS and science subgroup participants' responses to the item "I can explain how I reach the correct answer on a math test" were significantly correlated with their MCRT (HHS: $\rho = .440, p < .01$, science: $\rho = .549, p < .01$) and CRT (HHS: $\rho = .365, p < .05$, science: $\rho = .479, p < .01$). Fo the item, "I am usually confident I will do well on math tests" both of these subgroups' responses were significantly positively correlated with their CRT scores (HHS: $\rho = .420, p < .05$, science: $\rho = .339, p < .05$). Finally, HHS subgroup participants' responses to the item "I have experienced difficulties in math since high school or before high school" were significantly negatively correlated with their CRT scores ($\rho = -.377, p < .05$).

With respect to the other items of the SMPI several were also significantly correlated with MCRT and CRT scores. For participants with no major, their MCRT scores were significantly correlated with their responses items “I have a ‘game plan’ that is effective for tackling math homework” (ρ =.980, $p < .01$), “I am definitely a ‘work before play’ type of person” (ρ =.894, p<.05), and “I am doing better than I thought I would in college” ($\rho = .894, p < .05$). For participants in the business majors subgroup, their responses to the item “I expect to use the math I have learned in my future career” were significantly correlated with their MCRT and CRT scores ($\rho = .968, p < .05$), as were their responses to the item “I often play catch-up in my classes” ($\rho = -.889, p < .05$). For the HHS majors, their responses to the item “Sometimes, I am disappointed in my test results because I studied a great deal” were significantly negatively correlated with their MCRT scores ($\rho = -.344, p < .05$), and their responses to the item “My confidence in academic skills has increased this semester” were significantly positively correlated with their CRT scores ($\rho = .352, p < .05$). The responses of the participants in the science major subgroup for the following items were significantly correlated with their MCRT and/or CRT scores: “I use my time between classes productively (MCRT: $\rho = -.322, p < .05$, CRT: ρ =-$.297, p < .05$); “I expect to use the math I have learned in my future career” (MCRT: $\rho = .449, p < .01$, CRT: $\rho = .467, p < .01$); and “My grades in college math have influenced what degree I can pursue (MCRT: $\rho = -.299, p < .05$).

Lastly, there were only two items of all the SMPI items for which the responses of the participants of the mathematics course level subgroup were significantly correlated with either their MCRT or CRT scores, both of which are related to future orientation. Their responses to the item “I am taking the necessary steps to obtain the career I desire” were significantly negatively correlated with their MCRT scores ($\rho = -.444, p < .05$), and their responses to the item “I am definitely a ‘work before play’ type of person” were negatively correlated with their CRT scores (ρ =-$.436, p < .05$). This would indicate that higher levels of future orientation as measured by these items are associated with lower MCRT and CRT scores. However, this likely due to the fact that the vast majority of all participants in the mathematics course level subgroup agreed to these items (e.g., 80% of this group either agreed or strongly agreed to the item “I am taking the necessary steps to obtain the career I desire”) and had such high scores on both the MCRT and CRT.

Mathematics Courses Taken in Senior Year of High School Table 4.3.7 includes the respective Spearman correlation coefficients calculated for the responses to each of the items of the LBHEM Scale for the participants in each of the senior year high school mathematics course level subgroups and their respective MCRT and CRT scores.

The table follows the same structure as previous tables (Tables 4.3.5 and 4.3.6). Recall, that the sizes of the subgroups are 11 participants who reported not taking a mathematics class their senior year of high school, 20 who took an algebra-related course, 25 who took a generally-defined course, 32 who took a precalculus course, and 30 who took an advanced level course.

For the participants in the advanced level high school mathematics course level subgroup, their responses to only two of the LBHEM items were significantly correlated with their MCRT or CRT scores. These were the items “The wording of math problems confuses me” (MCRT: $\rho = -.591,$ p<.01, CRT: $\rho =$-.609, p<.01) and “Math and/or anything with numbers has been an obstacle to my academic success” (MCRT: $\rho = -.450, p < .05$, CRT: $\rho = -.444, p < .05$).

Table 4.3.7

Correlations Between MCRT and CRT Scores and SMPI Items Not from the LBHEM Scale Across Academic Major Subgroups

SMPI Item		MCRT Score	CRT Score
I'm just not good at math.	None	-.494	-.642*
	Algebra	-.166	-.392
	General	-.372	-.204
	Precalculus	-.468**	-.471**
	Advanced Courses	-.191	-.209
The wording of math problems confuses me.	None	-.418	-.594
	Algebra	-.200	-.308
	General	-.283	-.138
	Precalculus	-.359*	-.274
	Advanced Courses	-.591**	-.609**
Math and/or anything with numbers has been an obstacle to my academic success.	None	-.616*	-.784**
	Algebra	-.305	-.369
	General	-.255	-.234
	Precalculus	-.286	-.405*
	Advanced Courses	-.450*	-.444*
I have experienced difficulties in math since high school or before high school.	None	-.342	-.671*
	Algebra	-.368	-.253
	General	-.130	-.105
	Precalculus	-.176	-.256
	Advanced Courses	-.059	-.039
I am usually confident that I will do well on math tests.	None	.477	.644*
	Algebra	.528*	.351
	General	.466*	.402*
	Precalculus	.201	.408*
	Advanced Courses	.484*	.494*
I can explain how I reach the correct answer on a math test.	None	.424	.587
	Algebra	.546*	.242
	General	.623**	.527**
	Precalculus	.365*	.385*
	Advanced Courses	.363*	.296
I push aside math assignments and do them last.	None	-.044	-.181
	Algebra	-.641**	-.290
	General	-.093	-.107
	Precalculus	-.389*	-.401*
	Advanced Courses	-.008	.070

The scores of the precalculus level subgroup participants on the MCRT and CRT were significantly correlated with their responses to the items “I’m just not good at math” (MCRT: $\rho = -.468, p < .01$, CRT: $\rho = -.471, p < .05$), “The wording of math problems confuses me” (MCRT: $\rho = -.359, p < .05$), “Math and/or anything with numbers has been an obstacle to my academic success” (CRT: $\rho = -.405, p < .05$) “I can explain how I reach the correct answer on a math test” (MCRT: $\rho = .365, p < .05$, CRT: $\rho = .385, p < .05$), and “I push aside math assignments and do them last” (MCRT: $\rho = -.389, p < .05$, CRT: $\rho = -.401, p < .05$).

The precalculus level subgroup had the most significant correlations between their responses to LBHEM Scale items and their MCRT and CRT scores. The MCRT scores for the participants in the algebra course level subgroup were significant for three of the LBHEM items, “I am usually confident that I will do well on math tests” (MCRT: $\rho = .528, p < .05$) and “I can explain how I reach the correct answer on a math test” ($\rho = .546, p < .05$). It should be noted that only correlations between their responses on the MCRT were significant, which is likely due to the fact that 95% of the algebra course level subgroup did not earn any points on the CRT.

Participants’ scores on the MCRT and CRT form the general course level subgroup were significantly correlated with their responses for two items of the SMPI, which were “I can explain how I reach the correct answer on a math test” (MCRT: $\rho = .623$, p<.01, CRT: $\rho = .527, p < .01$) and “ I am usually confident that I will do well on math tests” (MCRT: $\rho = .466, p < .05$, CRT: $\rho = .402, p < .05$). Lastly, the responses from participants that reported not taking a mathematics class their senior year of high school on three SMPI items were significantly related to their MCRT and CRT scores; these were “Math and/or anything with numbers has been an obstacle to my academic success” (MCRT: $\rho = -.616, p < .01$, CRT: $\rho = -.784, p < .01$), “I’m just not good at math” ($\rho = -.642, p < .05$), and “I am usually confident that I will do will on math tests" ($CRT\ \rho = .644, p < .05$). It should be noted that for the first of these items, the correlation was the strongest observed correlation across all groups

With respect to the other items of the SMPI several were also significantly correlated with MCRT and CRT scores. Participants' scores on the MCRT and CRT in the precalculus level subgroup were significantly correlated with several other items of the SMPI, including "I expect to use the math I have learned in my future career" (MCRT: $\rho = .525, p <.01$, CRT: $\rho = .509, p < .01$), "Sometimes, I am disappointed in my test results because I studied a great deal" ($\rho = -.406, p < .05$, CRT: $\rho = -.528, p < .01$), "I have a fairly clear idea of what I need to study now in order to have the career I want" (MCRT: $\rho = .360, p < .05$). On two of the items related to future orientation and academic habits of mind, participants' responses from this group were associated with lower MCRT and CRT scores, which is a finding that has been seen in several different subgroups. These items were "Thinking about the future I want makes me do more now to get that future" (CRT: $\rho = -.350, p < .05$) and "I break down long-term assignments and class assignments and work on the mover time" (MCRT: $\rho = -.452, p < .01$, CRT: $\rho = -.485, p < .05$). These findings will be investigated further in future research (see Chapter 5 for a discussion of this). For the participants in the general level group, their MCRT and CRT scores were significantly correlated to their responses to the items "I often play catch-up in my classes" (MCRT: $\rho = -.421, p < .05$), and "I am always well-prepared for math class" (MCRT: $\rho = .406, p < .05$, CRT: $\rho = .411, p < .05$). Finally, for the advanced course level subgroup, their responses to the item "My confidence in academic skills has increased this semester" were significantly positively correlated with their MCRT ($\rho = .537, p < .05$) and CRT scores ($\rho = .556, p < .05$).

4.3.5 Analysis: Correlational Analyses of MCRT and Composite Scores

Recall that the scales of the SMPI were derived from a principal axis factoring analysis with a Promax rotation on all SMPI items (refer to Chapter 3 for the details of this analysis). Five reliable scales emerged, each of which comprise items that were included to measure certain psychoeducational facets of college students' educational profile. These scales are the Limiting Beliefs, Habits, and Experiences Related to Mathematics Scale (LBHEM); the Academic Habits of Mind Scale (AHM); the Future Orientation Scale (FO); the Academic Self-Concept (ASC) Scale; and the Hesitancy to Engage with Faculty (HEF) scale. Recall that all of the items within each individual scale were coded on the same 5-point Likert scale. However, for some items, lower values of the scale (1=strongly disagree, 2=agree) may algin with higher values on another item (4=agree, 5=strongly agree). For example, in the LBHEM factor, higher values for the item "I'm just not good at math" are associated with lower values for the item "I am usually confident I will do well on math tests" which each indicate less desirable or limiting outcomes. Therefore, in the development of the scales described below, some items needed to be reverse coded so that each item aligns with one another on the same measured 5-point scale. Refer to Table 4.2.10 for information about the items comprising each scale and the corresponding calculated reliability coefficient (Cronbach's α).

Also, recall that composite scores were created from the scales described above for the purposes of analyzing responses related to the psychoeducational facets described by each factor/scale. The composites were created by directly adding the responses for the items on each scale. Recall that each of the scales were coded in the same structure (some items needed to be reverse coded before inclusion on the scale). The following analyses were conducted to determine if any significant associations exist between the SMPI Scale Composite Scores and the

MCRT and CRT scores for participants in individual subgroups. Spearman correlation coefficients were therefore calculated between the composite scores for the participants of each subgroup and their respective MCRT and CRT scores.

4.3.6 Analysis: Correlational Analysis of MCRT and Composite Scores Across Entire Sample

To determine if the composite scores for each scale were significantly correlated with MCRT and CRT scores for participants in the entire sample, a Spearman correlation analysis was conducted. The composite scores for all 130 participants were analyzed in relation to their MCRT and CRT scores. The results of this analysis are given in Table 4.3.8 below. The rows of this table represent the SMPI Scale Composite Scores and the columns represent the MCRT and CRT scores. Therefore, in the cells that are the intersections of these rows and columns, the Spearman correlation coefficient that was calculated for the two groups is given. For example, the cell at the intersection of the row titled "AHM Scale Composite Score" and the column "MCRT Score" includes the value $-.059$; this value is the Spearman correlation coefficient calculated between all participants' AHM Scale Composite Scores and their MCRT scores. Correlation coefficients in the table indicated with a superscript * indicate significance at the $\alpha = .05$ level, and coefficients in the table indicated with a double superscript ** indicate significance at the $\alpha = .01$ level.

Table 4.3.8

Correlations Between SMPI Scale Composite Scores and MCRT and CRT Scores Across the Entire Sample

SMPI Item	MCRT Score	CRT Score
LBHEM Scale Composite Score	-.540**	-.549**
AHM Scale Composite Score	-.059	-.078
FO Scale Composite Score	-.023	-.089
ASC Scale Composite Score	-.026	.027
HEF Scale Composite Score	-.094	-.152

The only Scale Composite Score that was significantly correlated with all participants' MCRT and CRT scores was the LBHEM Composite. This is consistent with the correlational analysis on individual items of the SMPI. The correlations between the LBHEM Scale Composite Score and both the MCRT and CRT were moderate in strength, negative, and significant at the $\alpha = .01$ level. This suggests that higher levels of limiting characteristics related to mathematics that are part of students' psychoeducational profile are associated with lower MCRT and CRT scores, and this association is moderate in strength. Moreover, the results show that this is the only set of psychoeducational facets measured by the SMPI that are related to all participants' performance on the MCRT and CRT. It should also be noted that the correlations for the other Scale Composite Scores in relation to the MCRT and CRT are weak, and none other than the correlation between the HEF Scale Composite Score and CRT Scores ($\rho = -.152$) are stronger than $-.1$ or $.1$.

4.3.7 Analysis: Correlational Analysis of MCRT and Composite Scores Across Mathematics Course Level Subgroups

To determine if the composite scores for each scale were significantly correlated with MCRT and CRT scores for participants in each of the mathematics course level subgroups, a Spearman correlation analysis was conducted. The results of this analysis are given in Table 4.3.9. The structure of the table and information presented in Table 4.3.9 is similar to that of Table 4.3.8 above. However, in Table 4.3.9, the correlations are given for each of the mathematics course level subgroups. That is, the correlation between Composite Scores of each subgroup of participants and their MCRT and CRT scores are given, respectively.

Table 4.3.9

Correlations Between SMPI Scale Composite Scores and MCRT and CRT Scores Across Mathematics Course Level Subgroups

SMPI Item		MCRT Score	CRT Score
LBHEM Scale Composite Score	Developmental Level	-.603*	-.453
	General Level	-.273*	-.382*
	STEM Level	-.534**	-.541**
	Mathematics Level	-.104	-.104
AHM Scale Composite Score	Developmental Level	.103	.304
	General Level	-.041	-.075
	STEM Level	-.076	-.079
	Mathematics Level	-.276	-.276
FO Scale Composite Score	Developmental Level	-.168	-.051
	General Level	.275*	.157
	STEM Level	-.056	-.046
	Mathematics Level	-.183	-.183
ASC Scale Composite Score	Developmental Level	.144	.154
	General Level	.027	.233
	STEM Level	.127	.188
	Mathematics Level	-.112	-.112
HEF Scale Composite Score	Developmental Level	-.329	-.177
	General Level	-.189	-.154
	STEM Level	.125	-.012
	Mathematics Level	.058	.058

Significant correlations were observed between participants' composite scores in the developmental, general, and STEM level groups and their scores on the MCRT and/or CRT. Each of the correlations that were significant were negative (indicating that higher levels of limiting characteristics related to mathematics are associated with lower MCRT and CRT scores). For the participants in the developmental group, their LBHEM Scale Composite Scores were only significantly correlated with their MCRT scores. This is likely due to the fact that only one participant from this group earned a non-zero score on the CRT. For the general group, the correlations between both their scores on the MCRT and CRT were significantly correlated with their LBHEM Scale Composite Scores, with the strongest correlation being observed with the

CRT score.

The strongest correlations that were observed were the those between the STEM participants' scores on the MCRT and CRT and their LBHEM Scale Composite Scores. Both correlations were negative (indicating that higher levels of limiting characteristics related to mathematics are associated with lower MCRT and CRT scores), moderate in strength, and were significant at the $\alpha = .01$ level. The only group for which no significant correlations were observed was the mathematics course level subgroup. The correlations between their MCRT and CRT scores and the composite score were weak and negative. This may suggest that although limiting characteristics related to mathematics may be negatively associated with lower MCRT and CRT scores for this group, the relationship is far less strong than in other groups. This may also be due to the fact that the participants in the mathematics course level subgroup had the lowest LBHEM Scale Composite Scores (i.e., the lowest levels of limiting characteristics related to mathematics) than all of the other mathematics course level subgroups and the highest scores on both the MCRT and CRT than all subgroups as well.

Only one other significant correlation was observed in this analysis. The FO Scale Composite Scores for the participants in the general level group were significantly correlated with their MCRT scores. The correlation was positive, indicating that higher levels of future orientation are associated with better scores on the MCRT and CRT for this group. This finding is likely due to the fact the vast majority of participants in this group agreed to each of the FO Scale items (see Table 4.2.16) and the highest CRT and MCRT scores (see Table 4.1.9) overall.

4.3.8 Analysis: Correlational Analysis of MCRT and Composite Scores Across Other Subgroups

Gender Subgroups: Males and Females To determine if the composite scores for each scale were significantly correlated with MCRT and CRT scores for participants in each of the gender subgroups (males: $n = 49$, females: $n = 78$), a Spearman correlation analysis was conducted. For the male subgroup, the only significant correlations that were observed were between their MCRT and CRT scores and their LBHEM Scale Composite Scores. Males' LBHEM Scale Composite Scores were significantly related to their MCRT scores ($\rho = -.573, p < .01$) and their CRT scores ($\rho = -.534, p < .01$). Both of these correlations were negative (indicating that higher levels of limiting characteristics related to mathematics are associated with lower MCRT and CRT scores), moderate in strength, and significant at the $\alpha = .01$ level. No other significant correlations between males' SMPI Composite Scale Scores and their MCRT and CRT scores were observed.

For the female subgroup, females' LBHEM Scale Composite Scores were significantly correlated with their MCRT ($\rho = -.431, p < .01$) and CRT ($\rho = -.502, p < .01$) scores. Similar to what was observed above, both correlations were negative, moderate in strength and significant at the $\alpha = .01$ level. Females' CRT scores were also significantly correlated with their ASC Composite Scale score ($\rho = .242, p < .01$). This suggests that more desirable characteristics related to academic self-concept are associated with higher scores on the CRT. Other than these reported correlations, no other significant correlations between males' SMPI Composite Scale Scores and their MCRT and CRT scores were observed.

Academic Year Subgroups To determine if the composite scores for each scale were significantly correlated with MCRT and CRT scores for participants in each of the academic year subgroups (first-year students: $n = 46$, sophomores: $n = 26$, juniors: $n = 34$, seniors: $n = 22$), a Spearman correlation analysis was conducted. For each of the academic year subgroups, the only significant correlations that were observed were between their MCRT and CRT scores and their LBHEM Scale Composite Scores. All correlations were negative (indicating that higher levels of limiting characteristics related to mathematics are associated with lower MCRT and CRT scores), at least moderate in strength, and significant at the $\alpha = .01$ level. For first-year students, their LBHEM Scale Composite Scores were significantly correlated with their MCRT ($\rho = -.396,\ p < .01$) and CRT ($\rho = -.383, p < .01$). scores. Sophomores' LBHEM Scale Composite Scores were significantly correlated with their MCRT ($\rho = -.572,\ p < .01$) and CRT ($\rho = -.629, p < .01$). Junior participants' LBHEM Scale Composite Scores were significantly correlated with their MCRT ($\rho = -.514,\ p < .01$) and CRT ($\rho = -.542, p < .01$). Senior participants' LBHEM Scale Composite Scores were significantly correlated with their MCRT ($\rho = -.764,\ p < .01$) and CRT ($\rho = -.702, p < .01$). It should be noted that the strongest correlations observed in this analysis were those for the senior level group, which aligns with the findings on individual LHBHEM items.

Academic Major Subgroups To determine if the composite scores for each scale were significantly correlated with MCRT and CRT scores for participants in each of the academic major subgroups (participants with no major: $n = 6$, education majors: $n = 15$, business majors: $n = 5$, health and human service (HHS) majors: $n = 35$, science majors: $n = 47$, mathematics majors: $n = 22$), a Spearman correlation analysis was conducted. The LBHEM Scale Composite Scores for participants in the in the business, HHS, and science major subgroups were significantly correlated with their MCRT and CRT scores. In particular, the LBHEM Scale Composite Score for those in the business major subgroup were significantly related to their

MCRT ($\rho = -.913, p < .05$) and CRT ($\rho = -.913, p < .05$) scores (participants in this subgroup had equivalent MCRT and CRT scores). The LBHEM Scale Composite Score for those in the HHS major subgroup were significantly related to their MCRT ($\rho = -.442, p < .05$) and CRT ($\rho = -.437, p < .05$) scores. The LBHEM Scale Composite Score for those in the science major subgroup were significantly related to their MCRT ($\rho = -.452, p < .01$) and CRT ($\rho = -.489, p < .05$) scores.

The LBHEM Scale Composite scores for the participants in the mathematics major subgroup were not significantly related with LBHEM, which aligns with previous SMPI item analyses on this subgroup. However, their FO Scale Composite Scores were significantly correlated with their MCRT ($\rho = -.438, p < .05$) and their CRT ($\rho = -.435, p < .05$) scores. This finding would suggest that higher levels of future orientation are associated with lower scores on the MCRT and CRT for participants in this subgroup. This finding is likely due to the fact the vast majority of participants in this group agreed to each of the FO Scale items and high CRT and MCRT scores. For the participants reporting not having a major, their MCRT scores (analyses with CRT scores cannot be conducted since all participants in this group did not have a non-zero score on the CRT) were significantly negatively correlated with their ASC Scale Score ($\rho = -.880, p < .05$). This finding would suggest that higher levels of academic self-concept are associated with lower scores on the MCRT for students in this subgroup. This finding is likely due to the small sample size of this group coupled with their low scores on the MCRT and moderate levels of agreement to ASC items.

No significant correlations were observed between any of the SMPI Scale Composite Scores and the MCRT and CRT scores for the participants in the education major subgroup. However, the strongest correlation observed for this group, although not significant, was between their LBHEM Scale Composite Scores and their MCRT ($\rho = -.438$) and CRT ($\rho = -.436$) scores. For the HHS major subgroup, in addition to the significant correlations that were observed for the LBHEM Scale Composite Score, the FO Scale Composite Scores for participants in this subgroup were also significantly correlated with their MCRT ($\rho = .397, p < .05$) scores. This finding suggests that for participants in this subgroup, higher levels of future orientation are associated with higher scores on the MCRT.

Mathematics Courses Taken in Senior Year of High School To determine if the composite scores for each scale were significantly correlated with MCRT and CRT scores for participants in each of the high school mathematics course level subgroups (did not take mathematics course: $n = 11$, took an algebra-level course: $n = 20$, took a general-level course: $n = 25$, took a precalculus-level course: $n = 32$, took an advanced-level course: $n = 30$), a Spearman correlation analysis was conducted. The LBHEM Scale Composite Scores for participants in the in each of these subgroups were significantly correlated with their MCRT and/or CRT scores. Each of these correlations were negative (indicating that higher levels of limiting characteristics related to mathematics are associated with lower MCRT and CRT scores), moderate in strength, and significant (at least at the $\alpha = .05$ level). In particular, for those who reported not taking a math course their senior year of high school, their LBHEM Scale Composite Scores were significantly correlated with their MCRT ($\rho = -.710, p < .05$) scores. For participants in the algebra-level subgroup, their LBHEM Scale Composite Scores were significantly correlated with their MCRT ($\rho = -.552, p < .05$) scores. LBHEM Scale Composite Scores for participants in the general-level course subgroup their were significantly correlated with their MCRT ($\rho = -.422, p < .05$) scores. The fact that the LBHEM Scale Composite scores are only significantly correlated with MCRT scores for participants in these subgroups is likely due to the fact that the CRT scores for these participants were so low in general.

For the participants in the precalculus-level subgroup, their LBHEM Scale Composite Scores were significantly correlated with their MCRT ($\rho = -.442, p < .05$) and CRT ($\rho = -.525, p < .05$) scores. For the participants in the advanced-level subgroup, their LBHEM Scale Composite Scores were significantly correlated with their MCRT ($\rho = -.495, p < .05$) and CRT ($\rho = -.476, p < .05$) scores. The only additional correlation between any of the other SMPI Scale Composite Scores and MCRT or CRT scores for any of these subgroups was observed for the advanced level. The ASC Scale Composite scores for the participants in the advanced-level subgroup were significantly related to both their MCRT ($\rho = -.394, p < .05$) and CRT ($\rho = -.369, p < .05$) scores.

4.3.9 Analysis: Logistic Regression Analyses

To determine the predictive power of composite scores and subgroup membership on CRT performance, binary logistic regression analyses were run. Recall that logistic regression analyses determine the influence of a set of independent variables on a dichotomous outcome variable (Agresti, 1990). Here the dichotomous outcome under analysis is a categorical response variable that classifies participants according to whether they earned a zero score on the CRT (did not answer any of the problems correctly) or non-zero score (they answered at least one CRT problem correctly). Moreover, this dichotomous outcome variable separates the original CRT score variable (which consists of four levels of zero score, score of 1, score of 2 and score of 3) into a variable consisting of two levels: zero score and non-zero score. Further analysis of this variable was warranted due to the fact that the majority of all 130 participants in the study (87 or 66.9%) did not answer any of the CRT problems correctly, i.e., earning a zero score.

Additionally, the sizes of the individual score groups were small (18 or 13.8% scored a 1, 16 or 12.3% scored a 2, and 9 or 6.9% scored a 3), and further analysis of each individual score grouping would be limited in scope. Therefore, the creation of the dichotomous variable and its further analysis provided a way to analyze the larger CRT score subgroups that emerged.

Several sets of logistic regression analyses were run with the dichotomous CRT score variable as the dependent or response variables. Various predictor variables were analyzed, which included individual SMPI items (as categorical predictors), SMPI Scale Composite Scores, and categorical variables indicating subgroup membership. The categorical variables indicating subgroup membership were included in the form of dummy variables, which are variables consisting of only zeroes and ones that indicate subgroup membership. For example, consider the ordinal categorical variable indicating mathematics course level subgroup. This variable was originally coded on at four levels, the first being the developmental level, the second being the general level, the third being the STEM level, and the fourth being the mathematics level.

To ensure the variable is used as a categorical variable indicating subgroup membership and not an ordinal variable (which would imply that the differences between the codes or the groups are equivalent), four dummy variables can be created that indicate the same information. The four dichotomous dummy variables (consisting only of values of zero or one), one for each level, indicate subgroup membership. For example, the dummy variable for the developmental course level subgroup would indicate a participants' membership to the subgroup if their value in this variable was a one, and indicate that they do not belong to this subgroup if their value for the subgroup was a zero. Four such variables were created and used in logistic regression analyses to indicate subgroup membership.

Variables that were significantly or moderately correlated with one another were not included in the same model since this would violate the multicollinearity assumption of logistic regression (see Chapter 3 for a discussion of this assumption). With this restriction, as mentioned above, several sets of logistic regression analyses were run with various predictor or independent variables, including individual SMPI items (as categorical predictors), SMPI Scale Composite Scores, and categorical variables indicating subgroup membership. However, only three models emerged as statistically significant which included statistically significant predictors of the dichotomous response variable indicating zero or non-zero CRT score. The first two models consisted only of the dichotomous dummy variables indicating mathematics course level subgroup membership. The third model consisted only of the LBHEM Scale Composite Score. The predictors in each of these models could not be combined into a single model since the dummy variables (and the original ordinal categorical variable indicating mathematics course level subgroup membership) were significantly correlated with one another.

4.3.10 Analysis: Logistic Regression Analyzing Mathematics Course Level Subgroup Membership in Predicting Zero or Non-Zero CRT Scores

The first binary logistic regression analysis was conducted with the dichotomous variable indicating zero or non-zero CRT score as the dependent or response variable, and with three of the four dummy variables indicating subgroup membership to each of the mathematics course level subgroups (see IBM, 2013, p. 641, for a description of the SPSS algorithm used for this analysis). Note that only three of the four dummy variables are needed since a zero-value for each of these indicates membership to the fourth. In terms of the model, the membership to the fourth subgroup is indicated by the intercept. First, the model metrics and analytic measures of this logistic regression analysis are discussed. A Chi-Square (χ^2) test-statistic was calculated for the model with degrees of freedom 3 to determine if the model is significantly different from the null model (a model with no predictors). For this model, the Chi Square test-statistic was statistically significant [$\chi^2(3) = 37.019, p < .001$] indicating that the model was statistically significant and significantly different than the null model.

A Hosmer-Lemeshow goodness-of-fit test, which determines if the model suitably fits the data (Peng et al., 2002); if the result of this test is not statistically significant ($p > .05$) then the model suitably fits the data. The Hosmer-Lemeshow goodness-of-fit test was not statistically significant for this model [$\chi^2(2) = 0, p = 1.00$], which indicates that the model suitably fits the data. Finally, the Cox and Snell and Nagelkerke R^2 measures were calculated, which provide additional information about the model (larger values for these measures indicate better model fit) but cannot be interpreted directly. The Cox and Snell measure for this model was $R^2 = .248$ and the Nagelkerke measure for this model was $R^2 = .345$.

Table 4.3.10, contains the metrics of each of the predictor variables (the dummy variables indicating membership to different mathematics course level subgroups). In the table, the β values for each predictor variable are the coefficients of the variable in the logistic regression equation calculated for the model along with the standard error for each are given. The calculated Wald χ^2 value is also given for each predictor variable (for each predictor variable in this analysis, the degrees of freedom for this value is 1), which is used to determine whether the variable is a statistically significant predictor of the logit of the response variable, is also given along with the corresponding significance (or p-value). The significance values in the table indicated with a superscript * indicate significance at the $\alpha = .05$ level, and coefficients in the table indicated with a double superscript ** indicate significance at the $\alpha = .01$ level.

Lastly, the table includes the exponentiated β value, expressed as e^β which transforms the β coefficient into one that is interpretable in the context of the model. Recall, that the logistic regression equation consists of variables and corresponding β coefficients calculated to minimize the prediction of the logit of the probability P of belonging to the zero score group is divided by the probability of belonging to the non-zero score group $(1 - P)$. Thus, since these coefficients predict the logit, or $\ln\left(\frac{P}{1-P}\right)$, they must be exponentiated in order to be interpreted in terms of the context of the model, or, in this case, as the odds, holding all other variables constant, of obtaining a non-zero score on the CRT.

Table 4.3.10

Model Metrics of Predictor Variables in Logistic Regression Analysis of Mathematics and STEM Course Level Subgroup Membership Predicting Zero or Non-Zero CRT Scores

Predictor/Independent Variable Name	β	S.E.	Wald χ^2 (degrees of freedom 1)	Sig.	e^β
Membership to the Mathematics Course Level Subgroup	3.230	.653	24.467	<.001**	25.286
Membership to the STEM Course Level Subgroup	2.041	.501	16.597	<.001**	7.696
Constant (Intercept)	-2.132	.400	28.433	<.001**	.119

This logistic regression model predicts the zero or non-zero scores for all participants in this study using two dummy variables indicating membership to the mathematics and STEM course level subgroups. Each of the predictor variables and the constant (model intercept) were statistically significant at the $\alpha = .01$ level. One metric to consider in evaluating the model is the classification percentage of the model. For every participant, the model predicts, using the information provided by the independent variables, whether the participant earned a zero or non-zero score (the values of the dependent variable) on the CRT. Then, this prediction is compared with the actual known classification of each participant (earning a zero score or non-zero score). This yields a prediction or classification accuracy percentage (how many the model correctly classified in a particular group out of the total number of actual classifications in that group). This model correctly classified 94.3% of participants (82 of the 87) of who earned a zero score on the CRT. However, the model only correctly classified 34.9% of participants (28 of 33) who earned a non-zero score. Overall, the classification accuracy was 74.6%.

The only predictor variables in the model, both of which were significant at the $\alpha = .01$ level, indicated subgroup membership to either the STEM or mathematics course level subgroups. Interpreting the e^{β} values for each predictor provides an indication of the increase in odds of earning a non-zero CRT score simply by being a member of these subgroups. The e^{β} value for the mathematics course level subgroup dummy variable was $e^{\beta} = 25.286$, which indicates that, holding all other variables constant, membership to the mathematics course level subgroup increases one's odds over 25 times of earning a non-zero score on the CRT. The e^{β} value for the STEM course level subgroup dummy variable was $e^{\beta} = 7.696$, which idincates that, holding all other variables constant, membership to the STEM course level subgroup increases one's odds over 7 times of earning a non-zero score on the CRT. Taken together, these results indicate that those taking mathematics and STEM level mathematics courses are more likely to earn a non-zero score than their peers in lower course level subgroups.

Another regression analysis was conducted using different dummy variables indicating mathematics course level subgroup membership as predictor variables and the same dichotomous response variable indicating zero or non-zero CRT score. For this model, the Chi Square test-statistic was statistically significant [$\chi^2(2) = 32.698, p < .001$] indicating that the model was statistically significant and significantly different than the null model. The Hosmer-Lemeshow goodness-of-fit test was not statistically significant for this model [$\chi^2(2) = 0, p = 1.00$], which indicates that the model suitably fits the data. The Cox and Snell measure for this model was $R^2 = .222$ and the Nagelkerke measure for this model was $R^2 = .309$. The following table (Table 4.3.11) contains the metrics of each of the predictor variables (the dummy variables indicating membership to different mathematics course level subgroups). Table 4.3.11 below follows the same structure as the previous table, Table 4.3.10.

Table 4.3.11

Model Metrics of Predictor Variables in Logistic Regression Analysis of General and Developmental Course Level Subgroup Membership Predicting Zero or Non-Zero CRT Scores

Predictor/Independent Variable Name	β	S.E.	Wald χ^2 (degrees of freedom 1)	Sig.	e^{β}
Membership to the General Course Level Subgroup	-2.531	.501	22.066	<.001**	.095
Membership to the Developmental Course Level Subgroup	-2.554	1.079	5.606	.018*	.078
Constant (Intercept)	.251	.252	.995	.319	1.286

This logistic regression model predicts the zero or non-zero scores for all participants in this study using two dummy variables indicating membership to the general and developmental course level subgroups. Each of the predictor variables were statistically significant, however, the constant (model intercept) was not statistically significant at the $\alpha = .01$ level. Several elements of this model were less desirable than the last model. First, as just described, the intercept of the model is no longer significant. Next, the Cox & Snell and Nagelkerke R^2 values were smaller for this model than the previous. The e^{β} value for the general course level subgroup dummy variable is $e^{\beta} = .095$, which indicates that, holding all other variables constant, membership to the general level course level subgroup increases one's odds by .095 times. Similarly, the e^{β} value for the developmental course level subgroup dummy variable is $e^{\beta} = .078$, which indicates that, holding all other variables constant, membership to the general level course level subgroup increases one's odds by .078 times.

What is notable about this model is the classification accuracy, which reflects a substantial improvement to the previous model in classifying those who will earn a non-zero score. The overall classification accuracy is less than the previous model, however, at 73.1%. The classification accuracy of those who earned a zero score decreased to 67.8% of participants (the model correctly classifying 59 out of 87 participants earning a zero score). However, the classification accuracy of those who earned a non-zero score increased from 34.9% in the previous model to 83.7% in this model (correctly classifying 36 out of 43 participants who earned a non-zero score). Although the overall accuracy is less, the accuracies for each of the individual classifications only differ by 15.9%, whereas before they differed by 59.4%.

4.3.11 Analysis: Logistic Regression Analyzing LBHEM Scale Composite Score in Predicting Zero or Non-Zero CRT Scores

A regression analysis was conducted using the LBHEM Scale Composite Score as a predictor variable and the dichotomous variable indicating zero or non-zero CRT score as the response variable. For this model, the Chi Square test-statistic was statistically significant $[\chi^2(1) = 41.172, p < .001]$ indicating that the model was statistically significant and significantly different than the null model. The Hosmer-Lemeshow goodness-of-fit test was not statistically significant for this model $[\chi^2(8) = 5.388, p = .715]$, which indicates that the model suitably fits the data. The Cox and Snell measure for this model was $R^2 = .271$ and the Nagelkerke measure for this model was $R^2 = .378$. The following table (Table 4.3.12) contains the metrics of the only predictor variable (the LBHEM Scale Composite Score) and the constant (intercept). Table 4.3.12 below follows the same structure as the previous table, Table 4.3.11.

Table 4.3.12

Model Metrics of Predictor Variables in Logistic Regression Analysis of LBHEM Scale Composite Score Predicting Zero or Non-Zero CRT Scores

Predictor/Independent Variable Name	β	S.E.	Wald χ^2 (degrees of freedom 1)	Sig.	e^{β}
LBHEM Scale Composite Score	-.246	.047	26.867	<.001**	.782
Constant (Intercept)	3.899	.869	20.132	<.001**	49.348

This logistic regression model predicts the zero or non-zero scores for all participants in this study using only the variable of the LBHEM Scale composite score. This variable and the intercept were both statistically significant at the $\alpha = .01$ level. Several elements of this model were more desirable than each of the previous models. First, this model is more simplistic, only involving one statistically significant predictor variable, alongside a statistically significant intercept. Second, the model metrics of the Cox and Snell and Nagelkerke R^2 values are larger in this model than the previous two. Thirdly, the overall classification accuracy of this model is equivalent to that of the first model, but with substantial improvements in the prediction of non-zero score classifications. The overall classification accuracy for this model was 74.6%, with the model correctly classifying 83.9% of participants (73 out of 87) who earned a zero score on the CRT and 55.8% of participants (19 out of 43) who earned a non-zero score.

In considering both of the previous models, the current model maintains the highest overall classification accuracy, while also improving upon the disparity between the two types of classification accuracies. It is important to note that the LBHEM Scale Composite Score, which combines self-reported information regarding limiting characteristics related to mathematics predicts zero and non-zero score classifications at the same rate, and in some respects more effectively, than a mathematics course leveling structure that separates course levels by attained milestones in mathematics and previous mathematics learning. Furthermore, the LBHEM Scale Composite Score although being an indirect measure, provides similar information than direct academic information regarding participants.

Chapter 5: Discussion, Conclusion, and Recommendations

This chapter begins with an overview of the study before providing a brief discussion of the results according to each research question. After this, a brief concluding section is given. Finally, recommendations for future researchers are given within the context of how to improve on investigations of this kind.

5.1 Overview of the Study

The current study was undertaken to extend previous research and investigate the relationships between mathematical expertise, inhibitory control, (the ability to suppress an initial, immediate reaction to a stimulus and engage in deeper, more reflective thought) and facets of college students' psychoeducational profile. Previous research investigating the influence of mathematical expertise on mathematical tasks requiring inhibitory control have relied heavily on inherently limited measures of numeracy, especially in studies that involved the Cognitive Reflection Test (CRT). Other studies have investigated the influence of different measures of mathematical expertise—particularly, whether individuals pursued a career in mathematics or did not—on instruments other than the CRT. The current study analyzed the link between CRT performance and mathematical expertise as measured by the extent of mathematics learning in the classroom, which does not rely on a one-time, scale measure of numeracy and provides a more granulated measure than the stark difference between experts and non-experts in mathematics.

In the current study, a modified version of the CRT was administered to 130 undergraduate students of differing levels of mathematical expertise. The students were grouped into four subgroups according to the mathematics courses in which they were enrolled, each set of courses separated measurable milestones of mathematical proficiency and expertise (e.g., minimum grade requirements in prerequisite courses, placement exam scores, etc.). The subgroups were formed according to the academic program structure of the university under study. These were: introductory (commonly referred to as “remedial” or “developmental”) courses in mathematics, which do not satisfy the university requirements in mathematics; courses that satisfy the general education requirement for students not in the STEM (science, technology, engineering, or mathematics) majors; courses required for STEM majors; and courses for majors in mathematics. A descriptive and inferential statistical analysis of the results was undertaken.

The current study also extended previous research on college students’ learning and development, which has consistently demonstrated the importance of measuring psychoeducational variables in relation to collegiate academic success. Previous research has indicated that instruments measuring domain-general psychoeducational variables were not specific enough to measure individual psychoeducational differences with respect to college students’ conception of, approach to, and engagement with mathematics. Additionally, until now, these variables have not extensively been investigated in college students studying mathematics in college at different levels.

Therefore, in the current study, a psychoeducational survey instrument (the Short Mathematics Psychoeducational Inventory or SMPI) consisting of domain-general and mathematics-specific items was administered to the aforementioned students in the four subgroups of mathematics courses. A rigorous analysis of the instrument (which included factor and scale reliability analyses) and the responses to it (descriptive and inferential statistical analyses) was then undertaken.

Finally, the current study contributed new findings to extant literature related to the relationships among mathematical expertise measured in the way previously articulated, inhibitory control as measured by the CRT, and domain-general and mathematics-specific psychoeducational facets of college students' academic profile as measured by the SMPI. To analyze the relationships among these variables, descriptive and inferential statistical analyses of CRT scores and SMPI responses across each of the different mathematics course level subgroups was undertaken.

5.2 Discussion

5.2.1 Research Question 1: How do students from different collegiate mathematics courses perform on the Cognitive Reflection Test (modified by the investigator to include two additional problems)?

The performance of students from different collegiate mathematics courses on a modified version of the Cognitive Reflection Test (CRT) aligned with the hierarchy of the mathematics course level subgroups; that is, students in mathematics major courses performed significantly better than students at all other levels of mathematics courses; the students in the STEM course level subgroup exhibited the second highest performance; the third highest performance was exhibited by those in the general level subgroup, and the lowest performance was exhibited by those in the developmental level subgroup. Therefore, the results suggest that those with greater levels of mathematical expertise perform better on the CRT than their peers of lesser expertise, a finding that aligns with previous research. Although this may also align with common sense—those with more mathematical training perform better on an assessment consisting of mathematical word problems—the findings build upon previous research in providing a specific indication of how inhibitory control ability is developed.

First, the current study utilized a measure of mathematical expertise that was not an inherently limited, one-time assessment scale measure of numeracy, and was instead a measure intimately related to mathematics learning in the classroom. Second, this measure was more granulated than what was used in previous studies in that the extent of mathematics classroom learning between individuals in each of the subgroups was less stark than the difference between those pursuing a career in mathematics (e.g., working mathematicians, individuals majoring in mathematics, etc.) and those who are not. This is important because the results, contextualized through additional subgroup analyses (e.g., academic major, academic year, gender, mathematics courses taken in high school) that were conducted, provides an indication that individuals who spend more time learning mathematics in the classroom are able to better activate their inhibitory control abilities than their peers who have spent less time. Moreover, this provides evidence that inhibitory control ability, specifically as it is related to mathematics, is developed in mathematics classrooms rather than as a general cognitive ability common to academic learning in all disciplines, which can be accounted for by standardized measures of cognitive intelligence, which is what numerous previous studies have used to account for differences in inhibitory control abilities, especially as measured by the CRT.

Although it might seem obvious that those enrolled in higher level mathematics classes perform better on the CRT, what this study adds is a potential indication of why. Research has shown that the construction of mathematical habits, ways of thinking, and responses to mathematical stimuli are constructed socially in the students' mathematics classes (Schoenfeld, 1988). Therefore, based upon the results of the current study, it seems as if inhibitory control, a cognitive ability known to be related to success on mathematical tasks, is developed socially in students' mathematics classes. Although inhibitory control is part of the central executive, and is related to general academic success, the current study provides yet another indication of the specific relationship between inhibitory control and mathematics learning and doing, and that the mathematics classroom is an arena in which this ability is specifically developed.

The results of the current study also suggest that inhibitory control as such a cognitive ability is significantly underdeveloped in sample of college students under study. The majority of all participants (66.9%) did not correctly solve any of the CRT tasks requiring inhibitory control, which included participants in every mathematics course level subgroup. Additionally, the vast majority of all participants answered both of the decoy questions (which do not require inhibitory control to solve) correctly. This coupled with the fact that the majority of all incorrect answers were intuitive-incorrect (the intuitive, immediate response prompted by the language of the problem) indicates that the participants were engaging with the instrument intentionally, and that it was the nature of the CRT problems as tasks requiring inhibitory control that led to the incorrect answers.

Although the CRT problems are particularly designed to elicit a faulty response, and the environment in which they were administered may have contributed to individuals' performance, the problems themselves are nonetheless elementary in nature. Therefore, at the very least, these results suggest that inhibitory control, or the ability to "stop and think" (Vamvakoussi et al., 2012, p. 329), may need to be explicitly taught or reinforced as an essential skill related to mathematics learning and doing so that students may be better suited to engage with mathematical situations that specifically require it.

The results suggest that this skill is being developed in individuals who have taken more advanced mathematical courses both in high school and in college. This may be due to the fact that the requirement of having to stop and think about mathematical material increases as the mathematical content becomes more advanced. As one learns more mathematics, one also learns more heuristics (or general problem solving techniques) and more about the relationships between different mathematical concepts. It is known that these heuristics and relationships can prove problematic in certain mathematical situations, and when this is the case, inhibitory control is needed to ignore salient, intuitive, and seemingly relevant responses or heuristics that are actually inconsistent with task at hand, so that one can engage in more reflective thought (effectively “stopping and thinking”). Therefore, it is reasonable to suggest that as an individual learns more mathematics, spends more time in mathematical classes, is exposed to more advanced mathematical content, is given more opportunities to recognize the need for inhibitory control in different mathematical situations and practice utilizing it, the more mature and developed their inhibitory control abilities will be.

However, more research is certainly needed to further explore inhibitory control development as well as the nature of students’ engagement with mathematics problem solving tasks that specifically require it. In the case of the former, a more careful consideration of the complex set of sociocultural factors[2] influencing the development of mathematics knowledge and skills related to problem solving and inhibitory control ability should be undertaken in future research of this kind. In the case of the latter, research on Expectancy-Value Theory (Eccles et al., 1983) has indicated that “that individuals choose to engage in tasks and activities that have high value to them and at which they expect to succeed” (Lauermann, 2017, p. 1540). Therefore, it is possible that students’ performance on the MCRT, which was administered in a particular environment as part of this empirical study, was influenced by their valuation of the tasks in this situation and the perception of their ability to be successful with them. Therefore, future research should attempt to account for these factors, which may help further contextualize and interpret the findings of the current study.

[2] See Goos, 2014, and Lerman, 2000, 2001, for discussions sociocultural research applied to mathematics education.

5.2.2 Research Question 2: How do students from different collegiate mathematics courses respond to a psychoeducational survey measuring domain-general academic habits of mind, future orientation, and academic self-concept, and mathematics-specific academic habits of mind, beliefs, and past academic experiences?

In analyzing the responses to the psychoeducational survey instrument (the Short Mathematics Psychoeducational Inventory or SMPI) for students from different mathematics course level subgroups, the majority of items for which significant differences were found were related to mathematics. On every such item, which included every item from the Limiting Beliefs, Habits, and Experiences Related to Mathematics (LBHEM) Scale, the students from the mathematics course level subgroup had the most desirable responses. Recall that this scale measures behaviors, habits, and experiences that limit students in their conception of, approach to, and engagement with mathematics. Therefore, the composite score for this scale provides a summative measure of such characteristics. The mean responses for students in each course level subgroup for this composite score aligned with the hierarchy of the leveling structure itself; that is, those in the mathematics level had the lowest mean (and thus, the most desirable responses), the STEM level had the next lowest mean (and thus, the second most desirable responses), the general level had the next lowest mean (and thus, the third most desirable responses), and the developmental level had the highest mean (and thus, the least desirable responses). Those who were in the mathematics major and those who took an advanced mathematics course their senior year of high school also had more desirable responses on this scale than their peers. However, subgroups that did not differ in a specific way related to mathematics (e.g., gender subgroups, academic year subgroups) did not differ significantly with respect to their responses to these items.

Therefore, these findings, coupled with the fact that the LBHEM Scale is the most reliable and well-defined scale of all the SMPI Scales, demonstrate the utility of the of the LBHEM Scale in describing a particular psychoeducational facet of students' academic profiles related to mathematics. Additionally, subgroups differing with respect to mathematics did not differ significantly with respect to the domain-general psychoeducational facets that were measured, such as academic habits of mind, future orientation, and academic self-concept. Therefore, the results of the current study also suggest that these domain-general psychoeducational facets are not sufficient to describe the differences in students' academic profiles with respect to mathematics. The findings discussed here suggest the presence of a particular psychoeducational facet specific to mathematics, which is related to behaviors, habits, and experiences that limit students in their conception of, approach to, engagement with mathematics. Moreover, the findings suggest that there is a particular set of underlying characteristics that are intimately and inextricably related to college students' development in the discipline.

Additionally, the nature of the psychoeducational items utilized in this study to measure these attributes were developed from previous research that indicates that characteristics that are amenable to change are more consistent and meaningful predictors of overall academic success in college (Ben-Avie & Darrow, 2019). Accordingly, these items seem to measure developmental characteristics related to mathematics that are amenable to change after college enrollment. Further development of these items and the instrument as a whole is certainly needed, which should include an attempt to widen the scope of the current items to address other sociocultural variables known to influence mathematics learning and doing that were not accounted for in this study. However, the results of the current study indicate the potential for this instrument to aid in informing instructional practices and classroom interventions aimed at improving students' belief systems in mathematics, and therefore their learning and developmental experiences in the discipline.

It should be noted that the notion of such characteristics contributing to the development of students' belief systems that limit their success and experiences in mathematics is not at all new. Other researchers[3] (e.g., Schoenfeld 1985, 2013, 2016, and Gómez-Chacón 2000, 2014, 2017) have extensively studied this phenomenon. However, what the current study contributes to this body of research is not only another indication that a student's relationship with mathematics is particular and bound within their belief systems about themselves as learners and the discipline itself, but also a next step in how to measure this. Although more research in this area is certainly needed, the current study provides a clear indication of the LBHEM Scale's reliability and utility in measuring a particular set of underlying characteristics specific to mathematics that are intimately related to college students' development in the discipline.

5.2.3 Research Question 3: What is the relationship between the performance of students from different collegiate mathematics courses on the modified version of the Cognitive Reflection Test and their responses to a psychoeducational survey measuring domain-general academic habits of mind, future orientation, and academic self-concept, and mathematics-specific academic habits of mind, beliefs, and past academic experiences?

The scores on the CRT for students from different collegiate mathematics courses in this study were correlated with their responses to several items of the SMPI, mainly those related to mathematics. All 7 LBHEM Scale items were among the significantly correlated items. The responses to each of the positively-coded items of the LBHEM Scale (meaning higher levels of agreement are associated with more desirable outcomes) were significantly positively correlated

[3] In fact, Gómez-Chacón (2000) identified a brief set of "limiting" beliefs related to mathematics held by one of their research participants.

with MCRT scores; and the responses to each of the negatively-coded items of the LBHEM Scale (meaning higher levels of agreement are associated with less desirable outcomes) were significantly negatively correlated with MCRT scores. Moreover, the only Scale Composite Score that was significantly correlated with MCRT performance across the entire sample was the LBHEM Scale Composite Score. This score was negatively correlated (since the composite score is a summative measure of limiting characteristics related to mathematics) with MCRT scores and CRT subset scores with moderate strength.

With respect to the mathematics course level subgroups, the LBHEM Scale Composite Score was significantly negatively correlated with the MCRT scores of participants from each of the mathematics course level subgroups except the mathematics level. Additionally, the LBHEM Scale Composite Score was not significantly correlated with MCRT scores for students in the mathematics major subgroup. Recall that a consistent relationship was observed between students' enrollment in advanced mathematical courses, engagement with advanced mathematical content, and membership to a formal mathematical academic community and more desirable scores on the MCRT and more desirable responses on the LBHEM Scale. The findings for the third research question seem to strengthen the evidence for these findings—particularly that desirable characteristics in both respects are related to one another. It could potentially be the case that these students have more success on mathematical tasks and have advanced further in mathematics because they are not limited in the same ways as their peers with respect to mathematics. It could also be the case that these students have become less limited in mathematics over time as a result of progressing through more advanced mathematical content and finding success with it. It seems as if both of these interpretations are plausible and likely coexist in some form.

The study of Gómez-Chacón (2014) of secondary school students is presumably the only one to have investigated CRT performance in relation to mathematics belief systems, which found a relationship between positive beliefs in mathematics and CRT performance. The current study focused on college students and limiting characteristics of students' belief systems and found complimentary results to Gómez-Chacón's. However, the current study administered a different and arguably more sophisticated set of psychoeducational survey items measuring these characteristics, and, through the administration of domain-general psychoeducational items, helped isolate the relationship between limiting belief systems in mathematics and inhibitory control as measured by CRT performance. That is, the current study determined that the relationship between limiting characteristics related to mathematics was particular in nature, since none of the other scale composite scores (measuring domain-general psychoeducational characteristics) were related to MCRT scores. The current study also provides an important step forward in further identifying the particular nature of this relationship in college students. The current study also contributes a new instrument (and in particular, a new scale) to be used in future investigations of this kind.

Although more research in this area is certainly needed, several potential explanations exist for the current results. One is that the development of limiting characteristics in mathematics are related to the development of inhibitory control. As mentioned in the discussion of the first research question, the development of inhibitory control in mathematics is perhaps at least in part constructed socially, in students' everyday mathematics classes. Therefore, the relationship between the development of inhibitory control ability in this way could also related to the development, or lack of development, of limiting characteristics related to mathematics. For if mathematical knowledge and inhibitory control can be constructed socially, so too can limiting characteristics related to mathematics.

It could also be the case that the relationship is directional in nature, where limiting beliefs in mathematics affect the problem solving behaviors (in particular, the activation of inhibitory control) on the CRT, which would align with previous research on the influence of affective variables on problem solving (McLeod, 1988; Gómez-Chacón, 2000; 2014). Recall that the logistic regression model prediction zero and non-zero scores on the CRT that included as the only predictor the LBHEM Scale Composite Score was as effective in predicting students' zero and non-zero CRT scores as the best model using mathematics course level subgroup information. This suggests that information about college students' psychoeducational profile was as useful in predicting mathematics problem solving performance as direct measures of mathematical expertise, and aligns with previous research on college student learning and development (e.g., Ben-Avie & Darrow, 2019).

However, it is possible that this finding is confounded by the strong correlation between CRT scores and the LBHEM Scale; the relationship between the LBHEM Scale and individuals of different mathematics course level subgroups; and the large number of zero CRT scores observed in the data. Also, future research is certainly needed to further investigate additional factors influencing the relationship between inhibitory control task success and limiting beliefs in mathematics observed in the current study, such as other sociocultural variables and participants' valuation of the inhibitory control tasks and the perception of their ability to be successful with them. However, at the very least, the results of the current study may provide a modest indication of the relationship between students' belief systems in mathematics (particularly, limiting beliefs) and their problem solving performance on tasks requiring inhibitory control.

5.2.4 Contributions to the Functional Theoretical Framework

The results of the current study provide additional information about the relationships among the theoretical constructs illustrated in Figure 2.9.1. Recall, that this diagram provided a limited, yet functional framework depicting the relationships between each of the theoretical variables discussed in the review of literature. This diagram represents a narrow view of what was known about the relationship between these theoretical variables prior to the current study. Since it has been argued that the current study has contributed new knowledge to the literature, an updated diagram is given below to illustrate these contributions in relation to this framework. The original framework is given in Figure 5.3.1 and the updated framework is given in Figure 5.3.2.

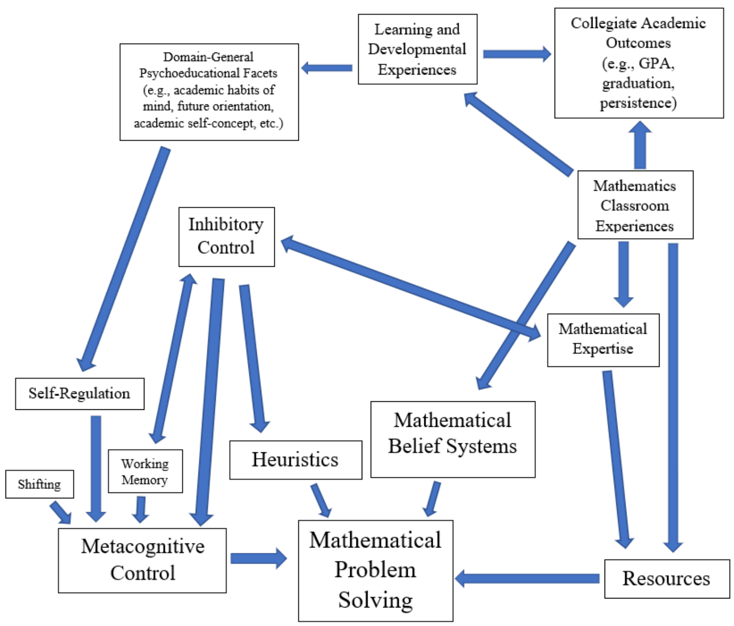

Figure 5.3.1: Functional Framework of Theoretical Constructs Discussed in the Review of Literature (Prior to the Current Study)

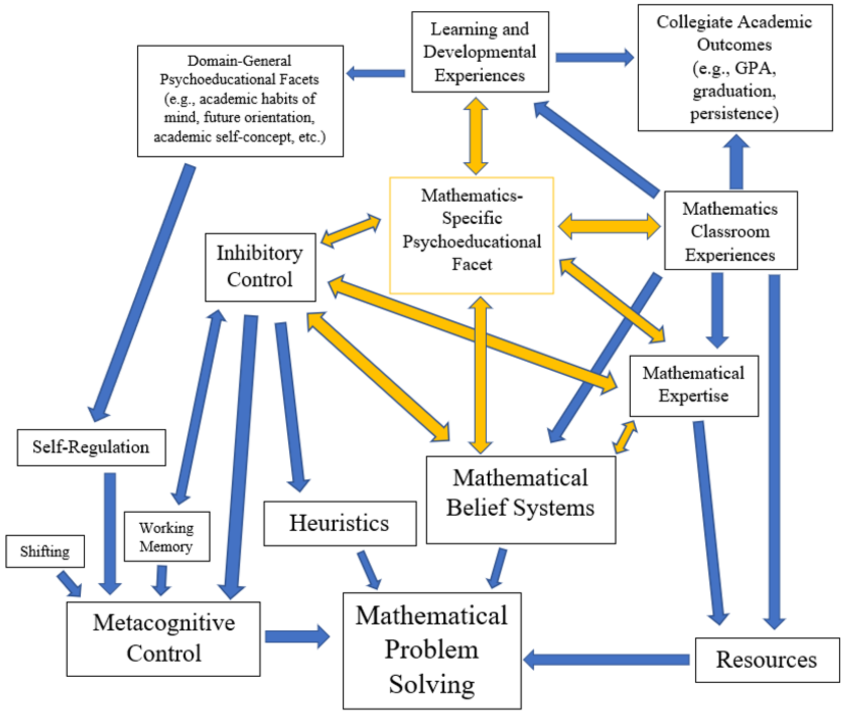

Figure 5.3.2: Functional Framework of Theoretical Constructs Discussed in the Review of Literature (After the Current Study)

Depicted in yellow in Figure 5.3.2 are the contributions of the current study to this functional framework. First, the presence the mathematics-specific psychoeducational facet that is distinct from the domain-general facets previously measured is illustrated. Not only do the results of the current study suggest a relationship between inhibitory control and this facet, they also suggest the relationship between this facet and other illustrated constructs. The nature of the items measuring indicators of this facet relate to students' belief systems in mathematics; their past and present developmental experiences, particularly in relation to the mathematics classroom; and the nature of their mathematical expertise. Therefore, the results suggest relationships between the mathematics-specific facet and these other constructs, including one between mathematics belief systems and mathematical expertise. Additionally, the current study reexamined the relationship between inhibitory control and mathematical expertise and seemed to support and strengthen findings from previous research establishing the existence of this relationship.

It is certainly the case that this framework was limited to begin with and that the findings of the current study provide only a modest and primary indication of the existence of many of these relationships. However, such an indication expands the theoretical work in the literature by providing new insight and avenues of future exploration. The current study inherently invites and calls for more research to strengthen our understanding of the relationships indicated here.

5.3 General Discussion and Conclusion

The current study was an exploratory empirical investigation of inhibitory control and the instrument used to measure it (the modified CRT); the responses to a new survey instrument measuring psychoeducational facets of college students' academic profile; and how these differ across different subgroups of college students enrolled in different levels of collegiate mathematics courses. This study was also innovative in measuring these variables in relation to one another. As such, it has contributed new knowledge to the fields of mathematics education and psychology new knowledge about college students' inhibitory control abilities and psychoeducational profile.

In totality, the results suggest that mathematical expertise is related to both performance on the CRT—and thus, inhibitory control abilities—and limiting psychoeducational characteristics related to mathematics, and that these are related to one another. These findings align with, and extend previous research on cognitive, psychological, and sociological variables related to problem solving and mathematics learning. The findings can potentially be explained by the nature of mathematics learning as inherently social and mathematics knowledge as a product of the everyday experiences in mathematics classrooms and communities.

Overall, the current study advances knowledge in the areas of inhibitory control and psychoeducational variables in the mathematics education of college students. The findings of this small-scale empirical study provide merely a modest step forward in these areas by providing another lens through which to view several phenomena already being extensively investigated by other researchers. Future work in these areas is certainly needed, not only to further investigate these phenomena, but also to determine if, and how such research can be applied to everyday mathematics instruction. For the aim of this work is to contribute additional knowledge regarding the complex set of variables influencing mathematics learning in the hopes that such knowledge can help improve the experiences and outcomes of mathematics students.

5.4 Recommendations for Future Research

In this section, recommendations for future research are given within the context of how the current study can be improved upon.

5.4.1 Sample Size and Recruitment

Recall that the current sample was relatively small and was selected by convenience, where willing faculty members in the mathematics department of the university under study created opportunities for the research to occur after their regularly-scheduled classes. Then, students from these classes were given the opportunity to voluntarily participate in the research. The nature of convenience sampling as a technique is inherently biased. Therefore, in future research of this kind, an attempt to incorporate randomization may help reduce this bias and potential influence on the results. Perhaps a larger set of individuals should be contacted initially so that from those voluntarily agreeing to participate, a random sample could be selected.

Also, it should be again noted that the sample consisted of participants from same institution in the northeastern United States, and therefore the results are not generalizable to other populations of college students from different institutions and from different regions of the United States or from different countries. Therefore, in future research of this kind, an attempt to recruit participants from different institutions should be made. The size of this sample was also relatively small, with only 130 participants. Although the purpose of the current study was exploratory in nature, the sample size is limiting in terms of developing a robust theoretical framework. The small sample also resulted in small and varying sizes of individual subgroups. For example, the size of the smallest course level subgroup was 11 (developmental level subgroup), followed 22 in the mathematics course level subgroup, 44 in the STEM level, and 55 in the general level. Therefore, any findings for the small subgroups, and across different subgroups should be considered with their sizes and the differences in these sizes in mind. Future research of this type should attempt to recruit greater number of participants in general, and more from the developmental and mathematics subgroups in particular.

In addition, in the current study, the participants in each course level subgroup were recruited from different mathematics courses. That is, the exact courses from which the participants were recruited differed greatly within each course level subgroup. For example, the participants from the general level subgroup were recruited from courses such as elementary statistics and mathematics for elementary school teachers, and in the STEM level subgroup, participants were recruited from courses such as calculus, calculus I, and intermediate statistics. The courses in each of these subgroups, respectively, differ greatly from one another. Although the nature of each of the course level subgroups is well-defined, the rigor and content covered in each of these classes vary. Therefore, in future research of this type, an attempt should be made to further account for these differences.

The current study was undertaken during the thirteenth, fourteenth, and fifteenth weeks of the semester at the university under study. Therefore, students who were not enrolled in mathematics courses at this time were not eligible to participate. The withdrawal deadline for the university under study occurs during the tenth week of classes, and thus, students who withdrew from the classes under study were not eligible to participate. The responses of these students to the SMPI and their scores on the MCRT may have provided more information about the relationship between limiting characteristics related to mathematics and the CRT. Students withdraw from mathematics courses for a variety of reasons, however one such reason is the difficulty and limitations encountered in the course. The SMPI could potentially measure such indicators and such data may have meaningfully contributed to the results of the study. Therefore, in future research of this type, data should be collected closer to the beginning of the semester to hopefully capture the perspective of students not present at the end of the semester.

5.4.2 SMPI Self-Reported Data and Question Format

One important limitation regarding the data collected on the SMPI is that all responses are self-reported. Information on all of the demographic and psychoeducational items is subject to the potential influence of inaccurate or misrepresented information. Future researchers should not that this potential limitation is relevant to all research involving survey instruments. As a result, inaccurate information collected for the psychoeducational items could have influenced the results gleaned from using the data on these items. Many items, including those on the LBHEM Scale, asked participants to report information about limiting characteristics related to mathematics, which may be personal and sensitive for some participants. Additionally, with respect to the demographic questions, participants were asked to report their gender, academic year, major, and information regarding whether they took a mathematics class their senior year of high school. It is possible that participants withheld or were not truthful regarding their membership to each of these subgroups.

There were also several demographic questions that were flawed, and as a result, the data gleaned from these items were omitted from analyses of different subgroups. One such question asked students about the how the mathematics course in which they were enrolled related to their degree requirements (whether the course was a prerequisite, satisfied their general education requirement, or was an elective). This question did not have a follow up question that provided participants with the opportunity to write in a response that differed from the options that were presented. Several participants wrote in such responses near the question itself, which indicated that the question did not account for all potential responses. Therefore, if this question is included in future administrations of the SMPI, more options should be provided, the wording of the question should be revised, and a box should be provided so that participants can write in additional responses

Another question asked participants whether they were placed into the course out of high school or whether they satisfied a prerequisite requirement to enter the course, or another reason. Unlike the previous question, this one did have an option for participants to write in their responses if they differed from the options that were presented. These write-in responses indicated that the question was not clear and as written, did not account for other circumstances related to enrollment and indicated that the data may be inaccurate as a result. If this question is to be included in future administrations of the SMPI, the wording of the question should be revised and more options should be provided in addition to the box for additional responses.

Another question asked participants whether they were taking the course in which they were enrolled for the first time. The original intent of the question was to further analyze course enrollment by investigating the scores or responses of individuals who are retaking the course in which they are enrolled. Additionally, previous research at the university has indicated that a sizeable number of individuals enrolled in certain courses under study retake the course. However, contrary to this, the vast majority of all participants indicated that they were taking the course for the first time. The number of individuals who reported retaking the course was small, and therefore, the data was not conducive to statistical analysis. In the future, the question should be reworded to be particular in nature, perhaps worded as “Are you retaking this class?” This may avoid potential responses that indicate that it is the first time they are taking the class with a particular professor. Another question that could be included in a future administration may ask whether or not the participants ever withdrew from a collegiate mathematics class. Such an item, like the ones previously described, may contribute additional information with which to analyze the experiences of individual students in collegiate mathematics classes.

In the future, additional subgroup questions should also be asked to further analyze both MCRT and SMPI data. Subgroup questions regarding gender were included since a great deal of previous research has analyzed the differences of CRT performance with respect to gender. However, it may be meaningful in the future to investigate differences with respect to race and ethnicity. The institution under analysis is diverse, and since the majority of all students at the university need to take mathematics classes part of their degree requirements, it was expected that the diversity of these classes would match the diversity of the university overall. Although this seemed to be the case, it would be important for future research to specifically account for this by including additional demographic questions on the survey instrument. Also, it may be the case that individuals from different racial or ethnic backgrounds may differ with respect to their responses to psychoeducational items measuring general and mathematics-specific academic habits of mind, future orientation, academic self-concept, and engagement with faculty.

Another issue related the self-reported data of the SMPI is related to the mathematics courses participants reported taking their senior year of high school. First, as noted above, there is no way to account for the potential misrepresentation of information on this subgroup question. Second, of the 118 participants that reported taking a math class their senior year of high school, 12 either did not answer the subsequent question asking them to specify the course that they took or could not recall (e.g., responses included “I don’t remember” or “I’m not sure”). Therefore, these participants were not included in the analyses conducted on the resulting subgroups. The information from these participants on both the MCRT and SMPI could have been meaningful in the analyses regarding this subgroup.

Third, for those who did report whether they took a high school mathematics class and indicated the class they took, the subgroups to which they belonged were potentially formed in a biased manner. Recall that the first group comprised the participants that did not take a math class their senior year of high school. The second group included participants that reported taking an algebraic course their senior year. The titles of these courses differed greatly, however, if the title included the word algebra or was known to primarily cover algebra (based upon regional knowledge of high school curriculum), it was included in this group. Course titles that were selected for this group included Algebra, Algebra II, Algebra III, Algebra and Statistics, ALEKS Program, College Algebra, Introduction to College Mathematics, Pre-Algebra, and Trigonometry and Algebra. However, there is no way to account for the rigor, content, and nature of each of these courses and likely vary greatly in these respects.

Recall that the next group included participants that reported taking courses in precalculus. Course titles that were selected for this group included precalculus, advanced algebra and precalculus, precalculus and trigonometry, and precalculus and statistics. Another group included participants that reported taking advanced courses. Course titles that were elected or this group included Calculus, AP Calculus, AP Statistics, and Multivariable Calculus. Again, there is no way to account for the rigor, content, and nature of each of these courses and likely vary greatly in these respects.

The final group included participants that reported taking courses with titles that were not aligned with the other groups and differed greatly from one another. This grouping also included courses are non-traditional when considering graduation requirements and trends in high schools in the northeast. These courses included accounting, general statistics, basic statistics, probability, and business mathematics. This is potentially the most problematic subgroup formation since the group does not have a well-defined structure like the others do. The courses that were included could be advanced in nature given their course titles and could have been grouped differently. There is likely great variation in the content covered, the rigor of each course, and how these relate to one another across the courses. Therefore, any results regarding this set of courses should be approached with knowledge of the potential bias inherent in the formation of the subgroup. In the future, a better accounting for these differences should be incorporated. Perhaps the collection of this information could be accomplished directly through academic profiles at the university. Additionally, there may exist a grouping of courses according to the university's admission requirement, which may help in reducing the bias associated with the primary researcher forming the groupings themselves.

5.4.3 Administration of MCRT before SMPI

It should be noted again that students took the MCRT prior to the SMPI, which could have influenced their responses to the SMPI. In essence, the taking of the CRT could have primed students to answer in particular ways, especially with respect to items that were related to mathematics. The experience of taking a mathematics assessment minutes before responding to a survey about mathematics beliefs may have influenced the nature with which students reported characteristics related to mathematics. Therefore, in a future study of this type, it would be advisable to administer the two instruments, the MCRT and the SMPI, at different times. This may aid in reducing the influence of taking the MCRT on SMPI responses. Perhaps the principal investigator could visit the courses to administer the MCRT or SMPI early in the semester, and then administer the second instrument later in the semester.

However, this would require a mechanism, presumably in the form of a primary data key, to link the responses between the two administrations. One of the reasons for the administration of the MCRT and SMPI together was to avoid the need for such a primary key, which would act as an individual identifier for students' responses and possibly dissuade potential participants form participating. In its current form, the study collected anonymous information from all student participants without any identifiers. Therefore, this would be an important consideration in a future administration.

5.4.4 A Focus on the CRT

It is important to note that the CRT was the sole instrument used to measure of inhibitory control in this study. The CRT was chosen as the instrument to measure inhibitory control in this study due to its demonstrated utility in measuring IC in adults through mathematical word problems. As noted in the literature review, the CRT has been characterized as "the dominant measure of adult individual differences in analytic versus intuitive thinking" (Young & Shtulman, 2020, p. 1396) or, in other words, differences in the ability to activate inhibitory control abilities to inhibit Type 1 responses and engage in Type 2 cognitive processes. The CRT has been administered extensively to adult participants, particularly college students since its creation (Frederick, 2005; Brañas-Garza et al., 2019). Additionally, unlike many instruments that have been developed to measure IC, the CRT does not rely on physical or spatial properties to produce a situation where inhibition is needed. Van Dooren & Inglis (2015) note that in contrast to such instruments, the CRT is "entirely cognitive: a salient response must be inhibited and some analytical work done before a correct answer can be given" (p. 714).

However, it should be noted that there are a large number of other instruments that have been extensively investigated in previous research to measure different aspects of inhibitory control. These instruments include general response inhibition and interference control tasks: modifications of the Stroop Task; numerical and non-numerical inhibition tasks; negative priming tasks; natural number bias tasks; proportional reasoning tasks; geometric reasoning tasks; probability tasks; and alternative forms of the CRT itself. Additionally, the CRT has been critiqued as an instrument simply measuring numerical and cognitive ability (e.g., Otero et al., 2022). There also exists evidence that suggests that the CRT has become less reliable due its popularity and strong correlations to measures of general intelligence (Haigh, 2016; Stieger & Reips, 2016; Otero et al., 2022; Welsh et al., 2013).

Therefore, it should be noted that the sole use of the CRT presents a potential limitation with respect to interpreting results related to inhibitory control. Although a large body of research suggests that the CRT is a legitimate measure of inhibitory control/cognitive reflection (Campitelli & Gerrans, 2014; Cokely & Kelley, 2009; Liberali et al., 2021; Toplak et al., 2011, 2014), is stable over time (Stagnaro et al., 2018), and that prior exposures to the instrument do not have a significant effect on future performance (Bialek & Pennycook, 2017; Stagnaro et al., 2018), the critiques of the instrument's use should be made clear. In future research of this type, additional and alternative instruments for measuring inhibitory control should be considered.

5.4.5 Engagement with the MCRT

It is important to note that the nature of students' performance on the MCRT in this study may be related to their engagement the instrument and how they perceive the activity of solving mathematical problems. Moreover, students' engagement with the MCRT was potentially influenced by the environment in which they took the assessment and the fact that they were participating in an empirical study. Previous research from Expectancy-Value Theory (EVT; Eccles et al., 1983) has demonstrated the importance of individuals' value of a task or activity and whether they expect to succeed on the task (Eccles et al., 1983, 2005; Eccles & Wigfield, 2020; Lauermann et al., 2017). In describing EVT, Lauermann et al. (2017) note that "the theory's basic premise is that individuals choose to engage in tasks and activities that have high value to them and at which they expect to succeed" (p. 1540).

In future research of this type, the potential influence of these factors on CRT performance should be accounted for. Perhaps if a survey instrument is administered alongside a version of the CRT, the survey should consist of items measuring students' value of, and expectations to be successful with mathematical problem solving tasks in general, and the CRT being administered in particular. This may help researchers determine the potential influence of these factors on MCRT performance and further contextualize any results gleaned from such measures.

5.4.6 Additional Outcome Measures

Following the discussion above in a previous section, the use of identifiers in future research may aid the investigation of additional outcome measures. In previous research of this kind, primary keys such as student ID numbers have been used. After IRB approval is sought and obtained for such use, the linkage of MCRT and SMPI data with other outcome measures such as grade point average (GPA), mathematics course grades, retention, and graduation may further contextualize the research findings of the current study. Although the use of additional identifiers may influence potential participants' decisions whether to participate, findings related to these measures may contribute to a more robust explanatory framework for findings of the current study.

5.4.7 Revisions to MCRT

In the current study, the MCRT consisted of the original CRT (designed by Frederick, 2005) along with two decoy questions (designed by Thomson & Oppenheimer, 2016). Since the Frederick's original (2005) study, a large number of studies have been conducted investigating the CRT in different forms and with different formats. The current study utilized the original CRT in an open-ended response format. Other studies have investigated the CRT with multiple-choice formats (Brañas-Garza et al., 2019; Sirota & Juanchich, 2018), alternate forms of the CRT (Thomson & Oppenheimer, 2016; Primi et al., 2015; Otero, 2019), and formats that relied on verbal responses rather than mathematical ones (Sirota et al., 2021). It would be worthwhile in future research to investigate the performance of students from different mathematics courses on different forms of the CRT. It would also be worthwhile to investigate performance on alternative versions of the CRT in relation to responses to the LBHEM Scale. Previous research has also investigated influence of problem ordering on CRT performance (Brañas-Garza et al., 2019). Therefore, in future administrations, it may also be worthwhile to investigate the potential influence of problem ordering, especially in the presence of decoy problems.

The follow up question that was included after each problem asking participants, “Have you seen this problem before?” should be revised. It seems as if the way in which this question was written invited participants to potentially overclaim their prior exposure to the problems. The question in its current form could be interpreted as asking “Have you seen a problem like this before?” instead of the intended question of “Have you seen this exact problem before?” Therefore, in a future administration, the latter should be given as the follow up question. This may help reduce the potential for participants to overclaim their prior exposure to the problems and provide a closer measure of whether they have seen the exact CRT problems before.

5.4.8 Revisions to SMPI

A previous section in this chapter mentioned the revision that is needed to the set of demographic questions of the SMPI. Also needed is the revision of psychoeducational items on the SMPI. Revisions are needed to each of the Scales of the SMPI to strengthen their reliability and comprehensiveness. Starting with the Academic Self-Concept (ASC) Scale, the current form of the scale does not meet conventional standards for a reliable scale since it consists of only two items. It is generally recommended for a scale to have at least three items, and therefore, additional items related to academic self-concept needed to be created and tested in future administrations. Next, the Future Orientation (FO) scale is comprised three items, each of which were responded to with exceptionally high levels of agreement. Although this is generally desirable since it indicates that a large number of student participants have high levels of future orientation, it limits the scale’s utility in measuring individual and subgroup differences with respect to future orientation. Both for the purposes of improving the scale and developing its utility to account for a greater variation in individual responses, new items need be created and tested through future administration.

Additional work on both the Academic Habits of Mind (AHM) and Hesitancy to Engage with Faculty (HEF) scales is also needed. In the case of the AHM Scale, the items of the scale were moderately correlated with items on the ASC and FO scales. One notable example of this was evident in the results of the factor analysis which revealed that the item "I am definitely a 'work before play' type of person" was significantly cross-loaded (in fact, with identical loadings) on both the AHM and FO Scales. Other items were also cross-loaded onto both the AHM and FO Scales, which indicates that further work is needed to make items on each scale more specific. However, this phenomenon is reasonable when interpreted within the context of both of the factors on which it loaded.

For example, consider the item "I am definitely a 'work before play' type of person" mentioned above. On the one hand, this item articulates an academic habit of mind consistent with desirable study skills. Those who "work before play" prioritize their work as more important than recreation, which is a desirable characteristic within the academic domain. On the other, those who do this have a mature future time perspective; they are able to understand that prioritizing their work is more important in the long-term than recreation. Thus, these individuals are oriented to the future, and thus, have desirable levels of future orientation, which is what the items comprising the third factor characterize. Although this interpretation is consistent with the structure of the factors and the theoretical foundation informing the creation of the SMPI, additional work may be needed to refine these scales to be more targeted in measuring the particular psychoeducational facets of future orientation. However, it could also be the case that future orientation and academic habits of mind are inextricably linked to one another, and measure a particular psychoeducational facet that is related to both. Either way, future research is needed to further refine these scales and investigate this phenomenon.

In the case of the HEF factor, additional items should be created and tested to further refine the scale. The results of the factor analysis conducted on SMPI items indicated that the items of the HEF factor were well-defined. These items emerged together on a single factor without any cross-loadings with items from other factors. Therefore, it appears that these items are measuring a distinct phenomenon not accounted for by the items of other scales. Additionally, previous research has indicated the importance of sense of belonging in the learning and development of college students (Ben-Avie & Darrow, 2019). It seems that the hesitancy to engage with faculty is related to sense of belonging in the classroom, and the comfortability one feels within the classroom among their peers. Investigating new items that measure HEF alongside those that measure sense of belonging may be a meaningful step in refining the HEF Scale.

The most reliable and well-defined scale of the SMPI was the LBHEM Scale, which measures limiting characteristics related to mathematics. The results of the current study seem to indicate that the LBHEM Scale measures a particular psychoeducational facet specific to mathematics, which is related to behaviors, habits, and experiences that limit students in their conception of, approach to, and engagement with mathematics. Moreover, the findings suggest that there is a particular set of underlying characteristics that are intimately and inextricably related to college students' development in the discipline that are not accounted for by scales measuring academic habits of mind, future orientation, academic self-concept, and hesitancy to engage with faculty. Therefore, more research is certainly needed to develop and refine the LBHEM Scale. Continued development of scale may help further characterize the nature of this psychoeducational facet related to mathematics that has emerged as significantly related to mathematical outcomes in the current study.

Recall from a previous section that performance on the MCRT was potentially influenced by the value students placed on solving the mathematical tasks and their expectations of being successful with them. It was recommended that in future research of this type, the potential influence of these factors should be accounted for, perhaps in the form of additional items on a survey instrument. Items of this nature may also help develop the LBHEM Scale and the SMPI in general. The inclusion of items that measure attributes related to expectancy and value may help further characterize individual's belief systems in mathematics as well as attributes that are amenable to change in this regard. Moreover, such items may help contextualize further the findings from other items, particularly those from the LBHEM Scale administered in this study.

Previous research has also addressed the complex set of sociocultural factors influencing mathematics learning (e.g., Goos, 2014; Lerman, 2000, 2001). As previously mentioned, these factors may have influenced both the nature of participants' performance on the MCRT and belief systems (particularly, limiting belief systems) in mathematics. Therefore, future research should attempt to carefully account for these factors. One way to accomplish this is to add items to the SMPI that measure attributes related to the sociocultural factors not accounted for in this study. The inclusion of such items may further characterize individuals' limiting belief systems and help further develop the LBHEM Scale and the SMPI as an instrument.

5.4.9 Factor Analysis

In the analysis of the SMPI as an instrument, a principal axis factoring analysis with a Promax rotation was conducted. This is an exploratory factor analytic technique, which aims to determine the underlying (or latent) structure of an instrument and group the measured variables that contribute to each of the latent constructs together (Fabrigar & Wegener, 2012). This method of analysis was chosen since the current study was the first to administer the SMPI (a new survey). It is the case that the items measuring domain general academic habits of mind, future orientation, and academic self-concept included on the SMPI have been developed through previous research (Ben-Avie & Darrow, 2019); however, they have never been administered alongside items measuring similar traits but specific to the domain of mathematics. Moreover, prior to the administration of the SMPI, there were no data to indicate how these items would interact or whether the inclusion of both of these sets would influence the latent variable structure observed in previous research. Therefore, an exploratory factor analysis was selected for this study.

However, there is an argument for the use of a confirmatory factor analytic procedure for the current data. When the particular construct and the items used to measure it are known (and have been developed through previous research), a confirmatory factor analysis (CFA) procedure can be run to determine if sample data is consistent with the measurement structure (Finch, 2020; Long, 1983). There are several confirmatory factor analytic techniques that can be used in an exploratory fashion by accounting for the known relationships among several variables in the measurement of a latent structure. Such an approach may be advantageous since it is known that several sets of the items of the SMPI are related to one another and were adapted from previously developed scales (see Ben-Avie et al., 2012 and Ben-Avie & Darrow, 2019). Therefore, a clear avenue of future research regarding the SMPI is to utilize confirmatory factor analytic techniques. Although SPSS can be limiting in terms of its use for this purpose, the open source statistical programming software R can be used for this. Additionally, several additional packages have been developed for the R software that can be used for confirmatory factor analyses, such as the lavaan package. In addition to the refinement and development of the SMPI as an instrument, these analyses will be explored further for the current set of SMPI data.

Also, recall that there is a great deal of disagreement in the literature regarding the appropriate sample size, number of variables, and the ratio between the number of variables and number of latent constructs for a factor analysis. Several empirical investigations have shown that commonly used rules for these values are inconsistent across studies, that model metrics provide more information about the appropriateness of an analysis, and that larger sample sizes do not necessarily imply better analyses (Arrindell & Van der Ende, 1985; de Winter & Dodou, 2012; Hogarty et al., 2005; MacCallum et al., 1999). However, other authors (e.g., Irwing & Hughes, 2009) have recommended sizes of at least 200 or greater for a factor analysis. In the current study, there were 130 participants, which falls short of recommended 200 minimum given by some authors.

The ratio between observed variables (31) to the number of participants (130) is greater than 1:4, which is within many of the standards given in previous literature (McCallum et al., 1999) and may provide a reasonable sample size for a factor analysis (de Winter & Dodou, 2012) when principal axis factoring is used and factor loadings (i.e., the correlational relationship between the factor and individual variables) are high. However, in general, the study's sample size and this ratio is far smaller than what is recommended by many authors, despite being within the standards of some. In future research on the SMPI, greater sample sizes will be sought so that the results of any factor analyses may be interpreted with greater strength and with caution than those in the current study.

The results of the factor analysis in the current study may be related to the choice of rotation, its parameters, and the implications these have on the interpretation of the factors.

Recall, that the Promax rotation, Fabrigar & Wegener (2012) note, begins with an orthogonal rotation (specifically, the Varimax rotation after a Kaiser normalization, see IBM, 2013, p. 292-293) and “then conducts a mathematical transformation of this initial solution, by raising factor loadings to a power of two or greater, to arrive at an oblique solution” (p. 78). The power to which the loadings are raised is known as kappa, and is most commonly used value for this is 4. Finch (2020) notes that “in practice, it is unusual for a researcher to alter kappa from the default value of 4” (p. 48). Therefore, consistent with previous research and accepted convention for psychological research of the type used in the current study, a Promax rotation with kappa equal to 4 was used. However, alterations to the kappa value may have influenced the results of the analysis. Therefore, further research and analysis regarding the appropriate kappa value for SMPI data is needed.

The output of a principal factor analysis with a Promax rotation includes a pattern matrix and a structure matrix. The pattern matrix, Fabrigar & Wegener (2012) note, contains the actual factor loadings and “are comparable to standardized partial regression coefficients” (p. 80). The structure matrix contains the correlations between individual items and the factors identified by the model. There exists some disagreement on how to appropriately interpret the factors using these two matrices. Generally, the literature suggest that researchers should use both matrices to interpret the nature of each of the factors (Gorsuch, 1983). However, recall that Fabrigar and Wegener (2012) note that “the pattern matrix should be the primary basis for interpreting factors” (p. 81), since this follows “the logic of oblique rotations” (p. 81). Indeed, the pattern matrix was used as the primary basis for interpreting the factors of the current study and developing the scales for the SMPI. However, it should be noted that items from the structure matrix were used in the development of SMPI scales. This may present a limitation in the interpretation of the factors and their description of an underlying latent structure. The LBHEM and HEF factors, however, were constructed using items only from the pattern matrix, consistent with the recommendations of Fabrigar and Wegener (2012) and others. This is encouraging since the LBHEM factor, in particular, was the most reliable scale and was involved in the most important findings of the current study.

With respect to the identification of factors, several potential limitations should be noted. Recall, that the Scree plot (Catell, 1966) should be analyzed, which plots the eigenvalues against the individual factors. In the plot, the points are connected, and one looks for the "last major drop" in these connections to indicate the number of factors that are defined by the model (Fabrigar & Wegener, 2012, p. 57). Analysis of the Scree plot in the current study indicated that the first four factors should be retained, with the fifth coming just after the "last major drop". Therefore, based upon conventions for interpreting the Scree plot, only four factors should have been retained, identified, and interpreted. However, in the current study the fifth factor (the HEF factor) was analyzed further due to its structure and its straightforward interpretability in context. The HEF factor was involved in several important findings in the current study. Therefore, these findings should be approached with the knowledge that the development of the HEF scale from the HEF factor relied upon its identification in the factor analysis.

Additionally, alternative methods of determining the number of factors may also have contributed to a different result. Parallel analysis is a recommended procedure that may offer a "more objective criterion" (Fabrigar & Wegener, 2012, p. 58) than what is obtained by the Scree test and the Kaiser criterion. This procedure, Fabrigar & Wegener (2012) note, "is based on the comparison of eigenvalues obtained from sample data with eigenvalues obtained from completely random data" (p. 58). Parallel analysis was briefly explored in the current study and is a clear direction for future analyses of the SMPI data, especially in different administrations of a refined version of the survey instrument. Other methods of rotation will also be explored in future work. One type of rotation which has gained recent attention is the Geomin rotation utilized in the Mplus software package. Since this is an oblique rotation similar to the Promax rotation, it may meaningfully contribute to the analysis of the SMPI data.

5.4.10 Logistic Regression

It is important to note that the logistic regression analyses in the current study were limited in scope. Only three models were analyzed and two of these only consisted of dichotomous ("dummy") variables indicating membership to one of the mathematics course level subgroups. Although the model helped further characterize the relationship between subgroup membership and CRT score, these models were simplistic and limited in their interpretation. The other model consisted of the LBHEM Scale Composite score as the only predictor. The simplicity of these models was a consequence of the multicollinearity of other potential predictor variables with these. Therefore, in future research, additional predictor variables will be considered in relation to zero and non-zero CRT scores.

Another limitation of the analysis is the lack of a well-defined cut-off value for each model. Each of the three models in the current study utilized the same cutoff value of .5 which is used in the regression model as a threshold probability for determining how to classify individuals in one of the two outcome categories (in this case, zero and non-zero CRT score). A cutoff value of .5 indirectly assumes that the general probability of belonging to one of the outcome groups is 50%. However, it is the case in the current study that the majority of all participants (66.9%) earned a zero score on the CRT. Therefore, since the cutoff value is below that of the actual proportion of membership to one of the outcomes of the dependent variable, the results may be skewed in this direction.

Recall that one of the metrics of the logistic regression model is the classification accuracy, i.e., the number of individuals correctly classified by the model into the dichotomous outcome group to which they actually belong. The potential skewing of the results by the cutoff value may influence this classification accuracy. That is, it could be the case that more individuals were correctly classified as being in the non-zero group simply because there are more individuals in that group to begin with. Therefore, future work will involve determining an appropriate cutoff value for the data of the current study. It should be noted that all three of the models in the current study utilized a cutoff value of .5, and therefore, each model may be skewed in the same manner. Although comparing different logistic regression models is not straightforward, one common metric for this is the classification accuracy. Thus, in comparing the different models, it should be noted that each may be skewed by the decision of the cutoff value. Since such comparisons helped contextualize certain findings in the current study, they should be approached with the knowledge of the potential influence of the cutoff value of .5. In future research, not only will each of the models be refined and improved, each model will be reassessed from a methodological standpoint to reduce bias and the potential skewing of results.

References

Allan, N. P., Hume, L. E., Allan, D. M., Farrington, A. L., & Lonigan, C. J. (2014). Relations between inhibitory control and the development of academic skills in preschool and kindergarten: A meta-analysis. *Developmental Psychology, 50*(10), 2368-2379.

Arrindell, W. A. & van der Ende, J. (1985). An empirical test of the utility of the observations-to-variables ratio in factor and components analysis. *Applied Psychological Measurement*, *9*(2), 165-178.

Attridge, N. & Inglis, M. (2015). Increasing cognitive inhibition with a difficult prior task: Implications for mathematical thinking. *ZDM Mathematics Education, 47*(5), 723-734.

Avram, L-A. (2018). Gender differences and other findings on the cognitive reflection test. *Studia Universitatis Babes-Bolyai, 63*(3), 56-67.

Babai, R., Shalev, E., & Stavy, R. (2015). A warning intervention improves students' ability to overcome intuitive interference. *ZDM Mathematics Education, 47*(5), 735-745.

Babai, R., Zilber, H., Stavy, R., & Tirosh, D. (2010). The effect of intervention on accuracy of students' responses and reaction times to geometry problems. *International Journal of Mathematics and Science Education, 8*(1), 185–201.

Barbera, J., Naibert, N., Komperda, R., & Pentecost, T. C. (2021). Clarity on Cronbach's alpha use. *Journal of Chemical Education*, *98*(2), 257-258.

Barkley, R. A. (1997). *ADHD and the nature of self-control.* New York: Guilford Press.

Baron, J., Scott, S., Fincher, K., & Metz, S. E. (2015). Why does the cognitive reflection test (sometimes) predict utilitarian moral judgement (and other things)? *Journal of Applied Research in Memory and Cognition, 4*(3), 265-284.

Ben-Avie, M. (2013) Do they stay or do they leave? *Technical Report of the Office of Assessment and Planning*, Southern Connecticut State University.

Ben-Avie, M. (2018). Focus on that which is amenable to change: EIA lessons from Southern Connecticut State University. *Assessment Update, 30*(1), 8-9.

Ben-Avie, M. (2008) Underprepared for the business sector: Seven design principles for interventions. *Journal of Business Management and Change*, 3(1), 7-34.

Ben-Avie, M., & Darrow, B., Jr. (2018). Using longitudinal cohort datasets. *Intersection: Quarterly Publication of the Association for the Assessment of Learning in Higher Education*, 3-5.

Ben-Avie, M., & Darrow, B., Jr. (2019). Malleable and immutable student characteristics: Incoming profiles and experiences on campus. *The Journal of Assessment and Institutional Effectiveness, 8*(1-2), 22-48.

Ben-Avie, M. & Polka, J. (2006, October 9). *From orientation to junior year: A psychological understanding of student engagement in higher education* [Conference presentation]. National Survey of Student Engagement (NSSE) Regional Workshop, New Haven, CT.

Ben-Avie, M., Haynes, N. M., Ensign, J., & Steinfeld, T. R. (2003a). Social and emotional development in relation to math and science learning—An introduction to the argument. In Haynes, N. M., Ben-Avie, M., & Ensign, J. (Eds.), *How social and emotional development add up: Getting results in math and science education* (pp. 1-8). Teachers College Press.

Ben-Avie, M., Haynes, N. M., White, J., Ensign, J., Steinfeld, T. R., Sartin, L. D., & Squires, D. A. (2003b). Youth development and student learning in math and science. In Haynes, N. M., Ben-Avie, M., & Ensign, J. (Eds.), *How social and emotional development add up: Getting results in math and science education* (pp. 9-35). Teachers College Press.

Ben-Avie, M., Kennedy, M., Unson, C., Li, J., Riccardi, R., Mugno, R. (2012) First-year experience: A comparison study. *Journal of Assessment and Institutional Effectiveness, 2*(2), 143-170.

Bialek, M. & Pennycook, G. (2017). The cognitive reflection test is robust to multiple exposures. *Behavioral Research Methods, 50*(5), 1953-1959.

Binet, A. (1903). *L'étude expérimentale de l'intelligence.* Schleicher Frères.

Blatt, S. J. & Stein, M. I. (1959). Efficiency in problem solving. *Journal of Psychology, 48*(2), 193-213.

Blair, C. & Razza, R. P. (2007). Relating effortful control, executive function, and false-belief understanding to emerging math and literacy ability in kindergarten. *Child Development, 78*(2), 647-663.

Bland, J. & Altman, D. (1997). *Statistics notes: Cronbach's alpha.* BMJ

Bloom, B. S. & Broder, L. J. (1950). Problem-solving processes of college students; An exploratory investigation. *Supplementary Educational Monographs, No. 73.* University of Chicago Press.

Bong, M., & Skaalvik, E. M. (2003). Academic self-concept and self-efficacy. *Educational Psychology Review, 15*(1), 1-40.

Borden, V. M. H. (2005). Identifying and analyzing group differences. In Coughlin, M. A. (Ed.), *Intermediate/advanced statistics in institutional research* (pp. 132-168). Association for Institutional Research.

Borella, E., Carretti, B., & Pelgrina, S. (2010). The specific role of inhibition in reading comprehension in good and poor comprehenders. *Journal of Learning Disabilities, 43*(6), 541‐552.

Brañas-Garza, P., Kujal, P., & Lenkei, B. (2019). Cognitive reflection test: Whom, how, when. *Journal of Behavioral and Experimental Economics, 82,* 101455.

Brown, F.E. and Corbin, J.N. (2004). Child development is the foundation of education. In E.T. Joyner, J.P. Comer, M.D., and M. Ben-Avie (Eds.). *Six pathways to healthy child development and academic success*. Thousand Oaks, CA: Corwin Press.

Brookman-Byrne, A., Mareschal, D., Tolmie, A. K., Dumontheil, I. (2018). Inhibitory control and counterintuitive science and maths reasoning in adolescence. *PLoS One, 13*(6), e0198973.

Bruner, J. S. (1962). *A study of thinking.* Science Editions.

Bruner, J. S. (1966). *Toward a theory of instruction.* Belknap Press.

Bull, R., & Lee, K. (2014). Executive functioning and mathematics achievement. *Child Development Perspectives, 8*(1), 36–41.

Buswell, G. T. (1956). *Patterns of thinking in solving problems.* University of California Publications in Education, University of California Press.

Byrne, B. M. & Shavelson, R. J. (1986). On the structure of adolescent self-concept. *Journal of Adolescent Psychology, 78*(6), 474-481.

Campitelli, G., & Gerrans, P. (2014). Does the cognitive reflection test measure cognitive reflection? A mathematical modeling approach. *Memory & Cognition*, *42*(3), 434-447.

Cattell, R. B. (1978). *The scientific use of factor analysis.* Plenum.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1*(2), 245-276.

Chiu, M. M. & Klassen, R. M. (2009). Relations of mathematics self-concept and its calibration with mathematics achievement: Cultural differences among fifteen-year-olds in 34 countries. *Learning and Instruction*, *20*(1), 2-17.

Clark, C. A., Pritchard, V. E., & Woodward, L. J. (2010). Preschool executive functioning abilities predict early mathematics achievement. *Developmental Psychology, 46*(5), 1176–1191.

Clement, J., Lockhead, J., & Monk, G. (1981). Translation difficulties in learning mathematics. *American Mathematical Monthly, 88*(4), 286–290.

Cokely, E. T., & Kelley, C. M. (2009). Cognitive abilities and superior decision making under risk: A protocol analysis and process model evaluation. *Judgement and Decision Making, 4*(1), 20-33.

Cole, J., Kennedy, M., & Ben-Avie, M. (2009). The role of precollege data in assessing and understanding student engagement in college. In R. M. Gonyea & G. D. Kuh (eds.), *Using NSSE in institutional research* (pp. 55-70). Jossey-Bass.

Comer, J. P. (2022, August 8). *Comer School Development Program: The Developmental Pathways Framework.* Yale School of Medicine, Child Study Center: Community Partnerships. https://medicine.yale.edu/childstudy/communitypartnerships/comer/pathways/

Comer, J. P. (1993). All children can learn: A developmental approach. *Holistic Education Review, 6*(1), 4-9.

Comer, J.P. (2005). Child and adolescent development: The critical missing focus in school reform. *Phi Delta Kappan*, *86*(10).

Comer, J. P. (1992). Child development: An organizing theme for school improvement. *Social Policy*, 28-30.

Comer, J.P., Ben-Avie, M., Haynes, N.M., Joyner, E.T. (1999). *Child by child: The Comer Process for change in education.* NY: Teachers College Press.

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

Coulanges, L., Abreu-Mendoza, R. A., Varma, Sashank, V., Uncapher, M. R., Gazzaley, A., Anguera, J., & Rosenberg-Lee, M. (2021). Linking inhibitory control to math achievement via comparison of conflicting decimal numbers. *Cognition 214, 104767*, 1-16.

Cox, D., R., & Snell, E., J. (1989). *The analysis of binary data* (2nd ed.). Chapman and Hall.

Cragg, L., Keeble, S., Richardson, S., Roome, H. E., & Gilmore, C. (2017). Direct and indirect influences of executive functions on mathematics achievement. *Cognition, 162*, 12-26.

Cragg, L. & Gilmore, C. (2014). Skills underlying mathematics: The role of executive function in the development of mathematics proficiency. *Trends in Neuroscience and Education*, *3,* 63-68.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297-334.

Davis, G. A. (1966). Current status of research and theory in human problem solving. *Psychological Bulletin, 66*(1), 36-54.

Darling-Hammond, L., Cook-Harvey, C. M., Flook, L., Gardner, M., & Melnick, H. (2018). *With the whole child in mind: Insights from the Comer School Development Program.* ASCD.

Darrow, B., Jr. (2020). *Moving beyond conversations and anecdotes: A pilot study of academic self-concept, word problem solving, and having an effective "game plan" in introductory college mathematics* [Unpublished master's thesis]. Teachers College, Columbia University.

Darrow, B. Jr. (2016). *On mathematics course placement, precollege and first-year data, and overall academic success in college: A longitudinal cohort study at a four-year public university* [Unpublished undergraduate thesis]. Southern Connecticut State University.

Deldoost, M. H., Mohammadzadeh, P., Saeedi, M. T., & Akbari, Akram (2019). The cognitive reflection test and numeracy as a predictor of students' choice of undergraduate major. *Journal of Educational, Cultural and Psychological Studies, 2019*(19), 147-162.

Dermitzaki, I., & Efklides, A. (2000). Aspects of self-concept and their relationship with language performance and verbal reasoning ability. *American Journal of Psychology, 113*(4), 621-638.

Desoete, A. & De Craene, B. (2019). Metacognition and mathematics education: An overview. *ZDM Mathematics Education, 51*(4), 565-575.

DeVellis, R. (2003). *Scale development: Theory and applications.* Thousand Oaks.

de Winter, J. C. F. & Dodou, D. (2012). Factor recovery by principal axis factoring and maximum likelihood factor analysis as a function of factor pattern and sample size. *Journal of Applied Statistics, 39*(4), 695-710.

De Weerdt, F., Desoete, A., & Roeyers, H. (2013). Behavioural inhibition in children with learning disabilities. *Research in Developmental Disabilities, 34*(6), 1998–2007.

Diamond, A. (2013). Executive functions. *Annual Review of Psychology, 64*(1), 135-168.

Dilts, R. B. (1999). *Sleight of mouth: The magic of conversational belief change*. Meta Publications.

Dineen, L. C. & Blakesley, B. C. (1973). Algorithm AS 62: Generator for the sampling distribution of the Mann-Whitney U statistic. *Applied Statistics, 22*(2), 269-273.

Dinno, A. (2015). Nonparametric pairwise multiple comparisons in independent groups using Dunn's test. *The Stata Journal, 15*(1), 292-300.

Dominowski, R. L. & Bourne, L. E. (1994). History of research on thinking and problem solving. In R. J. Sternberg (Ed.), *Thinking and problem solving: Volume 2 of handbook of perception and cognition* (pp. 1-35). Academic Press.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., Pagani, L. S., Feinstein, L., Engel, M., Brooks-Gunn, J., Sexton, H., Duckworth, K., & Japel, C. (2007). School readiness and later achievement. *Developmental Psychology, 43*(6), 1428–1446.

Duncker, K. (1945). On problem solving. *Psychological Monographs, The American Psychological Association, 58*(5), *(Whole No. 270)*.

Dunn, O. J. (1961). Multiple comparisons among means. *Journal of the American Statistical Association*, *56*(293), 52-64.

Dunn, O. J. (1961). Multiple comparisons using rank sums. *Technometrics, 6*(3), 241-252.

Dziuban, C. D. & Shirkey, E. C. (1974). When is a correlation matrix appropriate for factor analysis? *Psychological Bulletin, 81*(6), 358-361.

Eccles, J. S., Adler, T. F., Futterman, R., Goff. S. B., Kaczala, C. M., Meece, J. L., & Midgley, C. (19830. Expectancies, values, and academic behaviors. In J. T. Spence (Ed.), *Expectancies, values, and academic behaviors* (pp. 75-146). Freeman.

Eccles, J. S. (2005). Studying gender and ethnic differences in participation in math, physical science, and information technology. *New Directions for Child and Adolescent Development, 2005*(110), 7-14.

Eccles, J. S. & Wigfield, A. (2020). From expectancy-value theory to situated expectancy-value theory: A developmental, social, cognitive, and sociocultural perspective on motivation. *Contemporary Educational Psychology, 61,* 101859.

Evans, J. St. B. T. & Stanovich, K. E. (2013). Dual process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science, 8*(3), 223-241.

Fabrigar, L. R. & Wegener, D. T. (2012). *Exploratory factor analysis.* Oxford University Press.

Fagerlin, A., Zikmund-Fisher, B. J., Ubel, P. A., Jankovic, A., Derry, H. A., & Smith, D. M. (2007). Measuring numeracy without a math test: Development of the subjective numeracy scale. *Medical Decision Making, 27*(5), 672–680.

Finch, W. H. (2020). *Exploratory factor analysis.* SAGE Publications, Inc.

Fischbein, E. (1987). *Intuition in science and mathematics: An educational approach.* Kluwer Academic Publishers.

Frederick, S. (2005). Cognitive reflection and decision making. *The Journal of Economic Perspectives, 19*(4), 25-42.

Gagné, R. M. (1965). *The conditions of learning.* Holt, Rinehart, & Winston.

Gagné, R. M. (1966). Human problem solving: Internal and external events. In B. Kleinmuntz (Ed.) *Problem solving: Research, method, and theory* (pp. 128-148). Wiley.

Gathercole, S. E., Pickering, S. J., Knight, C., & Stegmann, Z. (2004). Working memory skills and educational attainment: evidence from National Curriculum assessments at 7 and 14 years of age. *Applied Cognitive Psychology,18*(1), 1-16.

Gómez-Chacón, I. M. (2000). Affective influences in the knowledge of mathematics. *Educational Studies in Matheamtics, 43*(2), 149-168.

Gómez-Chacón, I. M. (2017). Emotions and heuristics: The state of perplexity in mathematics. *ZDM Mathematics Education, 49*(3), 323-338.

Gómez-Chacón, I. M., García-Madruga, J. A., Vila, J. Ó., Elosúa, M. R., & Rodríguez, R. (2014). The dual processes hypothesis in mathematics performance: Beliefs, cognitive reflection, working memory and reasoning. *Learning and Individual Differences, 29,* 67–73.

Goos, M. (2014). Creating opportunities to learn in mathematics education: A sociocultural perspective. *Mathematics Education Research Journal, 26*(3), 439-457.

Gorsuch, R. L. (1983). *Fctor analysis* (2nd ed.) Erlbaum.

Gorsuch, R. L. (1980). Factor score reliabilities and domain validities. *Educational and Psychological Measurement, 40*(4), 895-897.

Gillard, E., Van Dooren, W., Schaeken, W., & Verschaffel, L. (2009a). Dual processes in the psychology of mathematics education and cognitive psychology. *Human Development, 52*(2), 95–108.

Gillard, E., Van Dooren, W., Schaeken, W., & Verschaffel, L. (2009b). Proportional reasoning as a heuristic-based process: Time constraint and dual task considerations. *Experimental Psychology, 56*(2), 92.

Gilmore, C., Keeble, S., Richardson, S., & Cragg, L. (2015). The role of cognitive inhibition in different components of arithmetic. *ZDM Mathematics Education, 47*(5), 771-782.

Hadamard, J. (1954). *The psychology of invention in the mathematical field.* Dover.

Haigh, M. (2016). Has the standard cognitive reflection test become a victim of its own success? *Advances in Cognitive Psychology, 12*(3), 145-149.

Hair, J., Black, W., Babin, B., Anderson, R., & Tatham, R. (2006). *Multivariate data analysis (6th ed.).* Prentice Hall.

Harter, S. (1985). *Manual for the self-perception profile for children*. University of Denver.

Henderson, K. B. & Pingry, R. E. (1953). Problem solving in mathematics. In H. F. Fehr (Ed.), *The learning of mathematics: Its theory and practice* (pp. 228-270). National Council of Teachers of Mathematics.

Hogarty, K. Y., Hines, C. V., Kromrey, J. D., Ferron, J. M., & Mumford, K. R. (2005). The quality of factor solutions in exploratory factor analysis: The influence of sample size, communality, and overdetermination. *Educational and Psychological Measurement*, *65*(2), 202-226.

Hosmer, D. W. & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). John Wiley & Sons, Inc.

Houdé, O., & Borst, G. (2014). Measuring inhibitory control in children and adults: Brain imaging and mental chronometry. *Frontiers in Psychology, 5*, 616-623.

Houdé, O., & Borst, G. (2015). Evidence for an inhibitory-control theory of the reasoning brain. *Frontiers in Human Neuroscience, 9*, 1-5.

Hudson, R. A., Kloosterman, P., Galindo, E. (2012). Assessing preservice teachers' beliefs about the teaching and learning of mathematics and science. *School Science and Mathematics, 112*(7), 433-442.

[IBM] International Business Machines Corporation. (2015). *IBM SPSS Statistics 22 algorithms.* IBM Corporation.

Jiang, R., Liu, R., Star, J., Zhen, R. Wang, J., Hong, W., Jiang, S., Sun, Y., & Fu, X. (2021). How mathematics anxiety affects students' inflexible perseverance in mathematics problem solving: Examining the mediating role of cognitive reflection. *The British Psychological Society, 91*(1), 237-260.

Jacob, R. & Parkinson, J. (2015). The potential for school-based interventions that target executive function to improve academic achievement: A review. *Review of Educational Research, 85*(4), 512-552.

Johnson, D. M. (1944). A modern account of problem solving. *Psychological Bulletin, 41*(4), 201-229.

John, E. R. (1957). Contributions to the study of the problem-solving process. *Psychological Monographs: General and Applied, 71*(8), *(Whole No. 447).*

Kahneman, H. (2011). *Thinking fast and slow.* Allen Lane.

Kaiser, H. F. & Rice, J. (1974). Little jiffy, mark IV. *Educational and Psychological Measurement, 34*(1), 111-117.

Kaiser, H. F. (1970). The application of electronic computers to factor analysis. *Educational and Psychological Measurement, 20*(1), 141-151.

Kaur, B. & Yeap, B. H. (2009). Mathematical problem solving in Singapore schools. In Kaur, B., Yeap, B. H., & Kapur, M. (Eds.), *Mathematical problem solving: Yearbook 2009, Association of Mathematics Educators* (pp.1-13). World Scientific.

Kelley, T. L. (1942). The reliability coefficient. *Psychometrika, 7*(2), 75-83.

Keren, G. & Schul, Y. (2009). Two is not always better than one: A critical evaluation of two-system theories. *Perspectives on Psychological Science, 4*(6), 533-550.

Kleinmuntz, B. (Ed.). (1966). *Problem solving: Research, method, and theory.* Wiley.

Kilpatrick, J. (2014a). History of research in mathematics education. In S. Lerman (Ed.), *Encyclopedia of mathematics education* (pp. 267-272). Springer.

Kilpatrick, J. (2014b). Mathematics education in the United States and Canada. In Karp, A. & Schubring, G. (Eds.), *Handbook on the History of Mathematics Education* (pp. 323-334). Springer.

Kilpatrick, J. (1969). Problem solving in mathematics. *Review of Educational Research, 39*(4), 523-534.

Kloosterman, P. & Stage, F. K. (1992). Measuring beliefs about mathematical problem solving. *School Science and Mathematics, 92*(3), 109-115.

Kompedra, R., Pentecost, T. C., Barbera, J. (2018). Moving beyond alpha: A primer on alternative sources of single-administration reliability evidence for quantitative chemistry education research. *Journal of Chemical Education, 95*(9), 1477-1491.

Kruglanski, A. W. & Gigerenzer, G. (2011). Intuitive and deliberative judgements are based on common principles. *Psychological Review, 118*(1), 97-109.

Kruskal, W. H. & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association, 47*(260), 583-621.

Krutetskii, V. A. (1969). An investigation of mathematical abilities in schoolchildren. In Jeremy Kilpatrick and Wirszup (Eds.), *Soviet studies in the psychology of learning and teaching mathematics: The structure of mathematical abilities: Volume 2* (pp. 5-57). School Mathematics Study Group, Stanford University.

Lakatos, I. (1976). *Proofs and refutations: The logic of mathematical discovery*. Cambridge University Press.

Lauermann, Tsai, Y-M., & Eccles, J. S. (2017). Math-related career aspirations and choices within Eccles et al.'s expectancy-value theory of achievement-related behaviors. *Developmental Psychology, 53*(8), 1540-1559.

Lazarte, M. E. (1933). *The development of problem-solving ability in arithmetic.* Clarke, Irwin, & Company.

Lerman, S. (2001). Cultural, discursive psychology: A sociocultural approach to studying the teaching and learning of mathematics. *Educational Studies in Mathematics, 46*(1/3), 87-113.

Lerman, S. (2000). The social turn in mathematics education research. In J. Boaler (Ed.), *Multiple perspectives on mathematics teaching and learning* (pp. 19-44). Ablex Publishing.

Leron, U., & Hazzan, O. (2006). The rationality debate: Application of cognitive psychology to mathematics education. *Educational Studies in Mathematics, 62*(2), 105–126.

Leron, U., & Hazzan, O. (2009). Intuitive vs analytical thinking: Four perspectives. *Educational Studies in Mathematics*, *71*(3), 263–278.

Liberali, J. M., Reyna, V. F., Furlan, S., Stein, L. M., & Pardo, S. T. (2012). Individual differences in numeracy and cognitive reflection, with implications for biases and fallacies in probability judgement. *Journal of Behavioral Decision Making, 25*(4), 361-381.

Lipkus, I. M., Samsa, G., & Rimer, B. K. (2001). General performance on a numeracy scale among highly educated samples. *Medical Decision Making, 21*(1), 37–44.

Long, J. S. (1983). *Confirmatory factor analysis.* SAGE Publications, Inc.

Lubin, A., Rossi, S., Lanoë, C., Vidal, J., Houdé, O. & Borst, G. (2016). Expertise, inhibitory control and arithmetic word problems: A negative priming study in mathematics experts. *Learning and Instruction, 5,* 40-48.

Lubin, A., Vidal, J., Lanoë, C., Houdé, O., & Borst, G. (2013). Inhibitory control is needed for the resolution of arithmetic word problems: A developmental priming study. *Journal of Educational Psychology, 105,* 701-708.

Maier, N. R. F. (1970). *Problem solving and creativity in individuals and groups.* Brooks/Cole Publishing Company.

Mann, H. B. & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically large than the other. *Annals of Mathematical Statistics, 18*(1), 50-60.

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods, 4*(1), 84-99.

Marsh, H. W., Hau, K. T., Balla, J. R., & Grayson, D. (1998). Is more ever too much? The number of indicators per factor in confirmatory factor analysis. *Multivariate Behavioral Research, 33*(2), 181-220.

McLeod, D. B. (1988). Affective issues in mathematical problem solving: Some theoretical considerations. *Journal for Research in Mathematics Education, 19*(2), 134-141.

Miller, M. B. (1995). Coefficient alpha: A basic introduction from the perspectives of classical test theory and structural equation modeling. *Structural Equation Modeling, 2*(3), 255-273.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41*(1), 49–100.

Morrison, F. J., Ponitz, C. C., McClelland, M. M. (2010). Self-regulation and academic achievement in the transition to school. In S. D. Calkins & M. Bell (Eds.), *Child development at the intersection of emotion and cognition* (pp. 203‑224). American Psychological Association.

Morsanyi, K., Busdraghi, C. Primi, C. (2014). Mathematical anxiety is linked to reduced cognitive reflection: A potential road from discomfort in the mathematics classroom to susceptibility to biases. *Behavioral and Brain Functions, 10*(31), 1-13.

Nagelkerke, N. J. D. (1991). A note on general definition of the coefficient of determination. *Biometrika*, *78*(3), 691-692.

[NCTM] National Council of Teachers of Mathematics (2000). *Principles and standards for school mathematics*. Lawrence Erlbaum, National Council of Teachers of Mathematics.

[NCTM] National Council of Teachers of Mathematics (2022). *Executive summary: Principles and standards for school mathematics.* Retrieved from https://www.nctm.org/standards/

Neill, W. T., Valdes, L. A., & Terry, K. M. (1995). Selective attention and inhibitory control of cognition. In Dempster, F. N. & Brainerd, C. J. (Eds.), *Interference and Inhibition in Cognition* (pp. 207-261). Academic Press.

Newell, A., Shaw, J. C., & Simon, H. A. (1958). Elements of a theory of human problem solving. *Psychological Review, 65*(3), 151-166.

Newstead, S. E. (2000). Are there two different kinds of thinking? *Behavioral and Brain Sciences, 23*(5), 690-691.

[NGA] National Governors Association: Center for Best Practices & Council of Chief State School Officers. (2010). *Common core state standards for mathematics.* Washington, DC: Authors.

Nunnally, J. & Bernstein, L. (1994). *Psychometric theory*. McGraw-Hill.

Obersteiner, A., Van Dooren, W., Van Hoof, J. & Verschaffel, L. (2013). The natural number bias and magnitude representation in fraction comparison by expert mathematicians. *Learning and Instruction, 28*, 64-72.

Obersteiner, A., Van Hoof, J., Verschaffel, L., & Van Dooren, W. (2016). Who can escape the natural number bias in rational number tasks? A study involving students and experts. *British Journal of Psychology, 107*(3), 537-555.

Osman, M. (2004). An evaluation of dual-process theories of reasoning. *Psychonomic Bulletin & Review, 11*(6), 988-1010.

Otero, I. (2019). *Validación de constructo y de criterio de la reflexibilidad cognitiva.* [Unpublished doctoral dissertation, Universidade de Santiago de Compostela].

Otero, I., Salgado, J. F., & Moscoso, S. (2022). Cognitive reflection, cognitive intelligence, and cognitive abilities: A meta-analysis. *Intelligence, 90*, 101614.

Ott, R. L. & Longnecker, M. (2001). An introduction to statistical methods and data analysis. Wadsworth Group.

Peng, C. Y. J., Lee, K. L., & Ingersoll, G. M. (2002). An introduction to logistic regression analysis and reporting. *The Journal of Educational Research, 96*(1), 3-14.

Pennycook, G., Cheyne, J. A., Koehler, D. J., & Fuglesang, J. A. (2016). Is the cognitive reflection test a measure of both reflection and intuition? *Behavioral Research Methods, 48*(1), 341-348.

Polya, G. (1945). *How to solve it.* Princeton University Press, Reprint 2014.

Polya, G. (1954). *Mathematics and plausible reasoning* (Vols. 1-2). Princeton University Press.

Polya, G. (1965). *Mathematical discovery* (Vols. 1-2). Wiley.

Polya, G. & Kilpatrick, J. (1974). *The Stanford mathematics problem book: With hints and solutions.* Dover Publications.

Primi, C., Morsanyi, K., Chiesi, F., Donati, M. A., & Hamilton, J. (2015). The development and testing of a new version of the cognitive test applying item response theory (IRT). *Journal of Behavioral Decision Making, 29*(5), 253-469.

Rimoldi, H. J. A., Fogliatto, H. M., Erdmann, J. B., & Donnelly, M. B. (1964). *Problem solving in high school and college students.* Loyola University Psychometric Lab.

Robinson, M. A. (2017). Using multi-item psychometric scales for esearch and practice in human resource management. *Human Resource Management, 57*(3), 739-750.

Rossi, S., Vidal, J., Letang, M., Houdé, O., & Borst, G. (2019). Adolescents and adults need inhibitory control to compare fractions. *Journal of Numerical Cognition, 5*(3), 314-336.

Schoenfeld, A. H. (1989). Explorations of students' mathematical beliefs and behavior. *Journal for Research in Mathematics Education, 20*(4), 339-355.

Schoenfeld, A. H. (1985). *Mathematical Problem Solving.* Academic Press, Reprint 2014.

Schoenfeld, A. (2016). Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics (Reprint). *The Journal of Education, 196*(2), 1-38.

Schoenfeld, A. H. (1988). Problem solving in context(s), In Randall I. Charles and Edward A. Silver (Ed.), *The teaching and assessing of mathematical problem solving* (pp. 82-92). Lawrence Erlbaum, National Council of Teachers of Mathematics.

Schoenfeld, A. (2013) Reflections on problem solving theory and practice. *The Mathematics Enthusiast, 10*(1 and 2), 9-34.

Schwartz, L. M., Woloshin, S., Black, W. C., & Welch, H. G. (1997). The role of numeracy in understanding the benefit of screening mammography. *Annals of Internal Medicine, 127*(11), 966–972.

Shtulman, A. & McCallum, K. (2014). Cognitive reflection predicts science understanding. *Proceedings of the Annual Meeting of the Cognitive Science Society, 36*(36), 2937-2942.

Simes, R. J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrtika, 73*(3), 751-754.

Siegel, S. (1956). *Nonparametric statistics for the behavioral sciences.* McGraw-Hill.

Sirota, M., Dewberry, C., Juanchich, M., Valuš, L., & Marshall, A. C. (2021). Measuring cognitive reflection without math: Development and validation of the verbal cognitive reflection test. *Journal of Behavioral Decision Making, 34*(3), 322-343.

Smith, J. P. (1973). *The effect of general versus specific heuristics in mathematical problem-solving tasks* (Order No. 7326637) [Doctoral dissertation, Teachers College, Columbia University]. ProQuest Dissertations.

Son, S-H. C., Choi, J. Y., & Kwon, K-A. (2019). Reciprocal associations between inhibitory control and early academic skills: Evidence from a nationally representative sample of head start children. *Early Education and Development, 30*(4), 456-477.

Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology, 15*(5), 72-101.

Spiegel, J. A., Goodrich, J. M., Morris, B. M., Osborne, C. M., & Lonigan, C. J. (2021). Relations between executive functions and academic outcomes in elementary school children: A meta-analysis. *Psychological Bulletin: American Psychological Association, 147*(4), 329-351.

Stanovich, K. E. & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences, 22*(5), 645-726.

Stavy, R., & Babai, R. (2010). Overcoming intuitive interference in mathematics: Insights from behavioral, brain imaging and intervention studies. *ZDM Mathematics Education, 42*(6), 621–633.

Stage, F. K. & Kloosterman, P. (1991). Relationships between ability, belief and achievement in remedial college mathematics classrooms. *Research and Teaching in Developmental Education, 8*(1), 27-36.

Stagnaro, M., Pennycook, G., & Rand, D. G. (2018). Performance on the cognitive reflection test is stable across time. *Judgment and Decision Making, 13*(3), 260–267.

Stieger, S. & Reips, U. D. (2016). A limitation of the cognitive reflection test: Familiarity. *PeerJ, 4,* e2395.

Stipek, D. (1998). *Motivation to learn.* Allyn and Bacon.

Stillman, G., & Mevarech, Z. (2010). Metacognition research in mathematics education: From hot topic to mature field. *ZDM Mathematics Education, 42*(2), 145–148.

Stoltzfus, J. C. (2011). Logistic regression: A brief primer. *Academic Emergency Medicine, 18*(10), 1099-10104.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643–662.

Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education, 48*(6), 1273-1296.

Tavakol, M. & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education, 2*(2), 53-55.

Taylor, J. M. G. (1987). Kendall's and Spearman's correlation coefficients in the presence of a blocking variable. *Biometrics, 42*(2), 409-416.

Thomson, K. S. & Oppenheimer, D. M. (2016). Investigating an alternate form of the cognitive reflection test. *Judgement and Decision Making*, *11*(1), 99-113.

Tipper, S. P. (1985). The negative priming effect: Inhibitory priming by ignored objects. *Quarterly Journal of Experimental Psychology*, *37*(4), 571–590.

Tipper, S. P. (2001). Does negative priming reflect inhibitory mechanisms? *Quarterly Journal of Experimental Psychology, 54*(2), 321–343.

Toplak, M. E., West, R. F., & Stanovich, K. E. (2014). Assessing miserly information processing: An expansion of the cognitive reflection test. *Thinking & Reasoning*, *20*(2), 147–168.

Toplak, M. E., West, R. F., & Stanovich, K. E. (2011). The cognitive reflection test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition, 39*(7), 1275–1289.

Vamvakoussi, X., Van Dooren, W., & Verschaffel, L. (2013). Educated adults are still affected by intuitions about the effect of arithmetical operations: Evidence from a reaction-time study. *Educational Studies in Mathematics, 82*(2), 323-330.

Van Dooren, W., & Inglis, M. (2015). Inhibitory control in mathematical thinking, learning and problem solving: A survey. *ZDM Mathematics Education, 47*(5), 713–721.

Van Dooren, W., De Bock, D., Hessels, An, Janssens, & Verschaffel, L. (2005). Not everything is proportional: Effects of age and problem type on propensities for overgeneralization. *Cognition and Instruction*, *23*(1), 57-86.

Van Hoof, J., Lijnen, T., Verschaffel, L., & Van Dooren, W. (2013). Are secondary school students still hampered by the natural number bias? A reaction time study on fraction comparison tasks. *Research in Mathematics Education, 15*(2), 154–164.

Verschaffel, L., Schukajlow, S., Star, J., & Van Dooren, W. (2020). Word problems in mathematics education: A survey. *ZDM Mathematics Education*, *52*(1), 1-16.

Wason, P. C. & Johnson-Laird, P. N. (1972). *Psychology of reasoning: Structure and content.* Harvard University Press.

Weller, J. A., Dieckmann, N. F., Tusler,M.,Mertz, C. K., Burns,W. J., & Peters, E. (2013). Development and testing of an abbreviated numeracy scale: A Rasch analysis approach. *Journal of Behavioral Decision Making, 26*(2), 198–212.

Welsh, M., Burns, N., & Delfabbro, P. (2013). The cognitive reflection test: How much more than numerical ability? *Proceedings of the Annual Meeting of the Cognitive Science Society, 35*(35), 1587-1592.

Wertheimer, M. (1945). *Productive thinking.* Harper.

Wickelgren, W. A. (1974). *How to solve problems: Elements of a theory of problems and problem solving.* W. H. Freeman.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin, 1*(6), 80-83.

Wilson, J. (1967). *Generality of heuristics as an instructional variable.* [Unpublished doctoral dissertation]. Stanford University.

Young, J. W. A. (1907). *The teaching of mathematics in the elementary and the secondary school.* Longmans, Green, & Company.

Young, A. G. & Shtulman, A. (2020). Children's cognitive reflection predicts conceptual understanding in science and mathematics. *Psychological Science, 31*(11), 1396-1408.

Yule, G. U. (1912). On the methods of measuring the association between two attributes. *Journal of the Royal Statistical Society, 75*(6), 579-652.

Appendix A

Modified Cognitive Reflection Test (MCRT) and Directions

Mathematics Problem Section

Please read:
On each of the following pages, there will be one mathematics problem for you to solve. After each problem, there will be a space to provide your final answer and place to indicate whether you have seen the problem before. Please write your final answer on the line provided and check the box (Yes/No) indicating whether you have seen this problem before.

You will have the entire rest of the page to use for your mathematical work. You may use only your pencil, the space provided, and a calculator to complete your work. The calculator must be a hand-held calculator—no phones, watches, or any other electronic devices are permitted for the duration of the testing period. If you do not have a pencil or calculator, please raise your hand and the researcher will provide you with these.

Problem 1:
A cargo hold of a ship had 500 crates of oranges. At the ship's first stop, 100 crates were unloaded. At the second stop, 200 more were unloaded. How many crates of oranges were left after the second stop?

Final Answer: ________crates

Have you seen this problem before? Yes ☐

No ☐

Problem 2:
A bat and a ball cost $1.10. The bat costs $1.00 more than the ball. How much does the ball cost?

Final Answer: $________

Have you seen this problem before? Yes ☐

No ☐

Problem 3:
If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets?

Final Answer: ________minutes

Have you seen this problem before? Yes ☐

No ☐

Problem 4:
An expedition on a mountain climbing trip was traveling with eleven horse packs. Each horse can carry only three packs. How many horses does the expedition need?

Final Answer: ________horses

Have you seen this problem before? Yes ☐

No ☐

Problem 5:
In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?

Final Answer: ________days

Have you seen this problem before? Yes ☐

No ☐

Appendix B

Short Mathematics Psychoeducational Inventory (SMPI)

Survey Section

Please read:

Recall that the purpose of this survey is to improve teaching practices in the mathematics classroom. None of your individual responses will be shared with anyone, everything you provide on this survey is completely confidential. No one will know anyone's individual responses since no identifying information of any kind will be collected here on the survey. Your Professor is not involved in the study and will not be reviewing any responses from anyone. Results of this study will be analyzed by the researchers and will only include summaries of all students' responses.

Please respond honestly to help improve mathematics education practices and ensure your academic experience is the best that it can be. Thank you very much for your participation and your help in continuing to improve the quality of education at the University.

1. My academic year is:
    a. Freshman/First-Year
    b. Sophomore
    c. Junior
    d. Senior
    e. Other:

2. My major is (please write in):

3. This math class I am taking:
    a. Is a prerequisite for another math class I need to take.
    b. Satisfies my LEP Quantitative Reasoning requirement.
    c. Is an elective for me.

4. This is the first time I am taking this class:
    a. Yes
    b. No

5. Please indicate the nature of your enrollment in this class:
    a. I placed into this class out of high school.
    b. I took successfully completed a prerequisite course requirement for this class.
    c. Other:

6. Did you take a math course in your senior year of high school?
    a. Yes
    b. No

7. If you took a math course your senior year of high school, which class did you take (please write in):

Please respond to each of the following by indicating your level of agreement with each statement. For each statement, please select one of the following: Strongly Disagree (SD), Disagree (D), Neutral (N), Agree (A), Strongly Agree (SA).

	SD	D	N	A	SA
I'm just not good at math.					
The wording of math problems confuses me.					
I tell my professor when I don't understand something from math class.					
The hardest part about solving word problems is understanding what is being asked.					
I use my time between classes productively.					
I settle for just passing my courses.					
I break down long-term assignments and/or class projects and work on them over time.					
I often play catch-up in my classes.					
I have a fairly clear idea of what I need to study now in order to have the career I want.					
Thinking about the future I want makes me do more now to get that future.					
I expect to use the math I have learned in my future career.					
Math and/or anything with numbers has been an obstacle to my academic success.					
I have experienced difficulties in math since high school or before high school.					
I am always well-prepared for math class.					
I am usually confident that I will do well on math tests.					
I wait until right before a math test to start studying.					
I have a 'game plan' that is effective for tackling math homework.					
I give myself enough time to really read course materials.					

I push aside math assignments and do them last.					
I am taking the necessary steps to obtain the career I desire.					
I can explain how I reach the correct answer on a math test.					
When something confuses me, I think about it until I can make sense out of it.					
I find it hard to prioritize my time.					
My confidence in academic skills has increased this semester.					
I study regularly to be successful in college.					
Sometimes, I am disappointed in my test results because I studied a great deal.					
Other commitments in my life get in the way of studying for classes.					
My grades in college math have influenced what degree I can pursue.					
I am hesitant to raise my hand in class even though I know the answer.					
I fear that if I ask for help, my professor will think less of me.					
I'm just not good at math.					
I am definitely a "work before play" type of person.					
I often get distracted during class by feelings of boredom.					
I am doing better than I thought I would in college.					

Is there anything else you would like us to know?

Appendix C

Additional Tables

Table 4.1.1

Entire Sample MCRT Performance by Number of Correct Answers

MCRT Performance	n	Relative Frequency Percentage	Cumulative Frequency	Cumulative Frequency Percentage
0 correct	1	.8	1	.8
1 correct	23	17.7	24	18.5
2 correct	66	50.8	90	69.2
3 correct	15	11.5	105	80.8
4 correct	16	12.3	121	93.1
5 correct	9	6.9	130	100

Note. The column of n provides the number of individuals with each level of performance. The relative frequency percentage was calculated using the total number of participants, 130.

Table 4.1.2

Entire Sample MCRT Performance by MCRT Problem

MCRT Problem	Correct		Incorrect	
	n	%	n	%
Problem 1	120	92.3	10	7.7
Problem 2	17	13.1	113	86.9
Problem 3	24	18.5	106	81.5
Problem 4	112	86.2	18	13.8
Problem 5	36	27.7	94	72.3

Note. The column of n provides the number of individuals either correctly or incorrectly answering each question. The % columns indicate the relative frequency percentage calculated using the total number of participants, 130.

Table 4.1.3

Entire Sample CRT Performance by Number of Correct Answers

CRT Performance	n	Relative Frequency Percentage	Cumulative Frequency Percentage
0 correct	87	66.9	66.9
1 correct	18	13.8	80.8
2 correct	16	12.3	93.1
3 correct	9	6.9	100

Note. The column of n provides the number of individuals with each level of performance. The relative frequency percentage was calculated using the total number of participants, 130.

Table 4.1.4

Entire Sample CRT Incorrect Answer Classification

CRT Problems	Total Incorrect	Intuitive-Incorrect		Non-Intuitive-Incorrect	
		n	%	n	%
MCRT Problem 2 ("ball and bat")	113	104	92	9	7.9
MCRT Problem 3 ("widgets")	106	87	82.1	19	17.9
MCRT Problem 5 ("lily pads")	94	78	83	16	17

Note. The columns indicated by n provide the number of individuals in each group and % columns indicate the relative frequency percentage calculated using the total number of participants in each group, indicated by the column "Total Incorrect".

For the tables on the following page (Tables 1.2.2 and 1.2.3), columns indicated by n provide the number of individuals within each mathematics course level subgroup with the indicated performance, and the % columns within each mathematics course level subgroup indicate the relative frequency percentage calculated using the total number of participants in each mathematics course level subgroup.

Table 4.1.9

Mathematics Course Level MCRT Performance by Number of Correct Answers

	Developmental Level (n = 11)			General Level (n = 55)			STEM Level (n = 44)			Mathematics Level (n = 20)		
MCRT Performance	n	Rel. Freq. %	Cum. Freq. %	n	Rel. Freq. %	Cum. Freq. %	n	Rel. Freq. %	Cum. Freq. %	n	Rel. Freq. %	Cum. Freq. %
0 correct	0	0	0	0	0	0	1	2.3	2.3	0	0	0
1 correct	6	54.5	54.5	9	16.4	16.4	8	18.2	20.5	0	0	0
2 correct	4	36.4	90.9	40	72.7	89.1	17	38.6	59.1	5	25	25
3 correct	1	9.1	100	4	7.3	96.4	8	18.2	77.3	2	10	35
4 correct	0	0	100	2	3.6	100	6	13.6	90.9	8	40	75
5 correct	0	0	100	0	0	100	4	9.1	100	5	25	100

Table 4.1.10

Mathematics Course Level MCRT Performance by MCRT Problem

MCRT Problem	Developmental Level (n = 11)				General Level (n = 55)				STEM Level (n = 44)				Mathematics Level (n = 20)			
	Correct		Incorrect		Correct		Incorrect		Correct		Incorrect		Correct		Incorrect	
	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%
Problem 1	8	72.7	3	27.3	54	98.2	1	1.8	38	86.4	6	13.6	20	100	0	0
Problem 2	0	0	11	100	1	1.8	54	98.2	7	15.9	37	84.1	9	45	11	55
Problem 3	0	0	11	100	2	3.6	53	96.4	12	27.3	32	72.7	10	50	10	50
Problem 4	8	72.7	3	27.3	47	85.5	8	14.5	37	84.1	7	15.9	20	100	0	0
Problem 5	1	9.1	10	90.9	5	9.1	50	90.9	16	36.4	28	63.6	14	70	6	30

Table 4.1.11

Mathematics Course Level CRT Performance by Number of Correct Answers

	Developmental Level ($n = 11$)			General Level ($n = 55$)			STEM Level ($n = 44$)			Mathematics Level ($n = 20$)		
CRT Performance	n	Rel. Freq. %	Cum. Freq. %	n	Rel. Freq. %	Cum. Freq. %	n	Rel. Freq. %	Cum. Freq. %	n	Rel. Freq. %	Cum. Freq. %
0 correct	10	90.9	90.9	49	89.1	89.1	23	52.3	52.3	5	25	25
1 correct	1	9.1	100	4	7.3	96.4	11	25	77.3	2	10	35
2 correct	0	0	0	2	3.6	100	6	13.6	90.9	8	40	75
3 correct	0	0	0	0	0	0	4	9.1	100	5	25	100

Note. The columns indicated by n provide the number of individuals within each mathematics course level subgroup with the indicated performance, and the "Rel. Freq. %" columns within each mathematics course level subgroup indicate the relative frequency percentage calculated using the total number of participants in each mathematics course level subgroup, and the "Cum. Freq. %" columns provide the sums of the relative frequency percentages up to and including the level of performance indicated.

Table 4.1.12

Mathematics Course Level CRT Incorrect Answer Classification

	Developmental Level ($n = 11$)			General Level ($n = 55$)			STEM Level ($n = 44$)			Mathematics Level ($n = 20$)		
CRT Problems	Total Inc.	Intuitive-Inc.		Total Inc.	Intuitive-Inc.		Total Inc.	Intuitive-Inc.		Total Inc.	Intuitive-Inc.	
		n	%		n	%		n	%		n	%
ball and bat	11	9	81.8	54	49	90.7	37	36	97.3	11	10	90.9
widgets	11	9	81.8	53	44	83	32	25	78.1	10	9	90
lily pads	10	5	50	53	44	83	28	24	85.7	6	4	66.7

Note. The total number of incorrect answers given to each problem within each mathematics course level subgroup is given in the column "Total Inc." column, and those that meet the criteria of being intuitive-incorrect answers are given in each respective column indicated by n. Each respective columns indicated by % provide the relative frequency percentages of each n calculated out of the total number of participants in each mathematics course level subgroup.

Table 4.2.11

Percentage Agreement and Disagreement on LBHEM Scale Items Across Mathematics Course Level Subgroups

SMPI Items	Mathematics Course Level Subgroups							
	Developmental ($n = 11$)		General ($n = 55$)		STEM ($n = 44$)		Mathematics ($n = 20$)	
	D%	A%	D%	A%	D%	A%	D%	A%
I'm just not good at math.	36.4	54.5	32.7	40	50	25	85	10
Math and/or anything with numbers has been an obstacle to my academic success.	36.4	54.5	38.2	47.3	59.1	20.5	90	5
I have experienced difficulties in math since high school or before high school.	36.4	63.6	21.8	69.1	50	38.6	70	25
I am usually confident that I will do well on math tests.	54.5	27.3	47.3	27.3	31.8	31.8	0	65
I can explain how I reach the correct answer on a math test.	9.1	36.4	14.5	58.2	6.8	75	5	95
I push aside math assignments and do them last.	45.5	45.5	40	40	54.5	20.5	80	5
The wording of math problems confuses me.	9.1	63.6	7.3	65.5	25	54.5	45	10

Note. The columns indicated by "D%" indicate the percentage of disagreement (including responses of disagree and strongly disagree) provided respectively by the participants of each mathematics course level subgroups to each item. The columns indicated by "A%" indicate the percentage of agreement (including responses of agree and strongly agree) provided respectively by the participants of each mathematics course level subgroups to each item.

Table 4.2.14

Percentage Agreement and Disagreement on AHM Scale Items Across Mathematics Course Level Subgroups

SMPI Items	Mathematics Course Level Subgroups							
	Developmental ($n = 11$)		General ($n = 55$)		STEM ($n = 44$)		Mathematics ($n = 20$)	
	D%	A%	D%	A%	D%	A%	D%	A%
I often play catch-up in my classes.	36.4	45.5	38.2	40	45.5	34.1	45	35
I wait until right before a math test to start studying.	36.4	18.2	50	18.5	38.6	38.6	50	30
I give myself enough time to really read course materials.	18.2	45.5	18.2	50.9	34.1	36.4	30	40
I settle for just passing my courses.	36.4	45.5	54.5	25.5	50	31.8	60	15
I break down long-term assignments and work on them over time.	36.4	45.5	36.4	41.8	36.4	36.4	45	30
I study regularly to be successful in college.	27.3	45.5	20	56.4	18.2	54.5	15	65
I find it hard to prioritize my time.	18.2	36.4	29.1	47.3	31.8	50	30	40

Note. The columns indicated by “D%” indicate the percentage of disagreement (including responses of disagree and strongly disagree) provided respectively by the participants of each mathematics course level subgroups to each item. The columns indicated by “A%” indicate the percentage of agreement (including responses of agree and strongly agree) provided respectively by the participants of each mathematics course level subgroups to each item.

Table 4.2.16

Descriptive Statistics of FO Scale Composite Score Across Entire Sample and Mathematics Course Level Subgroup

SMPI Items	Mathematics Course Level Subgroups							
	Developmental ($n = 11$)		General ($n = 55$)		STEM ($n = 44$)		Mathematics ($n = 20$)	
	D%	A%	D%	A%	D%	A%	D%	A%
I am taking the necessary steps to obtain the career I desire.	0	100	1.8	89.1	9.1	90.9	5	80
I have a fairly clear idea of what I need to study now in order to have the career I want.	36.4	45.5	7.3	72.7	4.5	75	5	70
Thinking about the future I want makes me do more now to get that future.	18.2	81.8	1.9	87	2.3	81.8	5	65

Note. The columns indicated by “D%” indicate the percentage of disagreement (including responses of disagree and strongly disagree) provided respectively by the participants of each mathematics course level subgroups to each item. The columns indicated by “A%” indicate the percentage of agreement (including responses of agree and strongly agree) provided respectively by the participants of each mathematics course level subgroups to each item.

Table 4.2.18

Descriptive Statistics of ASC Scale Composite Score Across Entire Sample and Mathematics Course Level Subgroup

SMPI Items	Mathematics Course Level Subgroups							
	Developmental ($n = 11$)		General ($n = 55$)		STEM ($n = 44$)		Mathematics ($n = 20$)	
	D%	A%	D%	A%	D%	A%	D%	A%
My confidence in academic skills has increased this semester.	36.4	63.6	21.8	43.6	29.5	45.5	20	40
I am doing better than I thought I would in college.	18.2	72.7	16.4	50.9	20.5	34.1	25	45

Note. The columns indicated by “D%” indicate the percentage of disagreement (including responses of disagree and strongly disagree) provided respectively by the participants of each mathematics course level subgroups to each item. The columns indicated by “A%” indicate the percentage of agreement (including responses of agree and strongly agree) provided respectively by the participants of each mathematics course level subgroups to each item.

Table 4.2.20

Descriptive Statistics of HEF Scale Composite Score Across Entire Sample and Mathematics Course Level Subgroup

SMPI Items	Mathematics Course Level Subgroups							
	Developmental ($n = 11$)		General ($n = 55$)		STEM ($n = 44$)		Mathematics ($n = 20$)	
	D%	A%	D%	A%	D%	A%	D%	A%
I fear that if I ask for help, my professor will think less of me.	81.8	9.1	47.3	34.5	65.9	18.2	80	15
I am hesitant to raise my hand in class even though I know the answer.	54.5	27.3	23.6	63.6	27.3	54.5	60	20
I tell my professor when I don't understand something from math class.	20	70	21.8	47.3	25.6	53.5	10	80

Note. The columns indicated by "D%" indicate the percentage of disagreement (including responses of disagree and strongly disagree) provided respectively by the participants of each mathematics course level subgroups to each item. The columns indicated by "A%" indicate the percentage of agreement (including responses of agree and strongly agree) provided respectively by the participants of each mathematics course level subgroups to each item.