RELATIONSHIPS BETWEEN GROUP-ADMINISTERED ABILITY TESTS AND INDIVIDUAL ACADEMIC ACHIEVEMENT

by

Elpiniki Marchesini
A Dissertation
Submitted to the
Graduate Faculty
of
George Mason University
in Partial Fulfillment of
The Requirements for the Degree
of
Doctor of Philosophy
Psychology

Committee:

_______________________ Director

Ellen W. Rowe _______________________

Tara M. Chaplin _______________________

AW _______________________ Department Chairperson

_______________________ Program Director

_______________________ Dean, College of Humanities and Social Sciences

Date: 5-7-2020 _______________________

Spring Semester 2020
George Mason University
Fairfax, VA

Relationships Between Group Administered Ability Tests and Individual Academic Achievement

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

by

Elpiniki Marchesini
Master of Arts
George Mason University, 2017
Bachelor of Science
Virginia Tech, 2015

Director: Timothy W. Curby, Professor
Department of Psychology

Spring Semester 2020
George Mason University
Fairfax, VA

DEDICATION

"So much of me is made of what I learned from you."

–Stephen Schwartz

Pappa, Mamma, Jacob and Lodovico

ACKNOWLEDGEMENTS

I would like to express my gratitude to the faculty and staff of the College of Humanities and Social Sciences and the Graduate School of Education at George Mason University. For the past five years, GMU has been my home. It's been an honor to have been surrounded by people so dedicated to students personal and intellectual development.

Specifically, I'd like to thank Dr. Ellen Rowe. Over the past five years, Ellen has served as my metaphorical salesperson, professor, supervisor, co-author, cheerleader, and friend. There are no words to fully express how much you have impacted my life. The way you care about your students does not go unnoticed, and I only hope my future students feel as supported as you've made me feel. I'll certainly miss walking past your office door at the end of the day (with your keys hanging) and popping my head in for a "quick" goodbye.

I'd also like to thank Megan Davis and the rest of GMU's Cognitive Assessment Program's staff. Megan, thank you for so graciously allowing my research endeavors to take up precious floor space in the CAP office (even though I'm sure at times it was a stressful sight to see) and for handling some of the more logistical aspects of making this thing happen! CAP Staff, I'm not quite sure how you never got sick of me. Thank you for the endless conversations and company. Namely, Lauren Vaughn, Kate Long, and Tiffany Lee for being the real MVPs and donating your time to this research in one way or another.

Last but not least, I'd like to thank Dr. Tim Curby. Three years ago, you provided me with the opportunity to pursue this doctorate degree, and for that I am eternally grateful. In this pursuit, you've forced me to challenge myself to think critically about topics I would have never in a million years though I'd be exploring. You also provided me my first opportunity to teach, which was the cherry-on-top to my Mason experience.

TABLE OF CONTENTS

Page
List of Tables viii
Abstract ix
Chapter 1: Conceptualizations of Giftedness 1
  Psychometric Models of Giftedness 1
  Systems Models of Giftedness 8
    WICS Model of Giftedness 8
    The Three Ring Conception of Giftedness 9
  Developmental Models of Giftedness 10
Chapter 2: Gifted Education and Identification 12
  The Gifted Child Paradigm for Gifted Education 13
  The Talent Development Paradigm for Gifted Education. 17
  The Differentiation Paradigm for Gifted Education 18
Chapter 3: Intelligence and Academic Achievement 21
  Reading 21
  Mathematics 24
Chapter 4: Group-Ability Tests and CHC Theory 28
Chapter 5: Research aims and hypotheses 33
Chapter 6: Method 37
  Participants 37
  Procedure 38
  Measures 39
    Group-Administered Cognitive Assessments 39
      The Naglieri Nonverbal Ability Test (NNAT) 39
      The Cognitive Abilities Test (CogAT) 41
    Academic Achievement 46
    Background History Questionnaire. 49

Review of Academic Records 50
Chapter 7: Data analysis 51
Chapter 8: Results 57
Broad Reading 59
Reading Comprehension 60
Broad Mathematics 60
Math Problem Solving 61
Math Calculation 61
Research Question #2 62
NNAT Score 63
Reading 63
Mathematics 64
CogAT Composite Score 65
Reading 65
Mathematics 65
NNAT vs. CogAT Composite 66
Research Question #3 69
CogAT Verbal Battery 69
Reading 69
Mathematics 70
CogAT Quantitative Battery 71
Reading 71
Mathematics 71
CogAT Nonverbal Battery 72
Reading 72
Mathematics 73
Battery Score Models 74
Research Question #4 77
Research Question #5 79
Discussion 81
Limitations 91
Implications 92

References.. 95
Biography.. 107

LIST OF TABLES

Table Page
Table 1 Descriptive Statistics of NNAT and CogAT Scores... 57
Table 2 Descriptive Statistics of Academic Achievement.. 58
Table 3 Correlations between NNAT and CogAT Scores.. 58
Table 4 Academic Achievement by Gifted Programming Group 59
Table 5 Reading Achievement Regressions 62
Table 6 Mathematics Achievement Regressions 63
Table 7 Stepwise Regressions NNAT and CogAT Composite (Reading Achievement) . 67
Table 8 Stepwise Regressions NNAT and CogAT Composite (Mathematics Achievement).. 68
Table 9 Stepwise Regressions Battery Score Models (Broad Reading, Basic Reading, Broad Mathematics).. 75
Table 10 Stepwise Regressions Battery Score Models (Math Problem Solving)............. 76
Table 11 Stepwise Regressions Battery Score Models (Math Calculation) 76
Table 12 Stepwise Regressions Nonverbal Measures (Reading Achievement) 77
Table 13 Stepwise Regressions Nonverbal Measures (Mathematics Achievement)........ 78

ABSTRACT

RELATIONSHIPS BETWEEN GROUP ADMINISTERED ABILITY TESTS AND INDIVIDUAL ACADEMIC ACHIEVEMENT

Elpiniki Marchesini, Ph.D.

George Mason University, 2020

Dissertation Director: Dr. Timothy W. Curby

Around six percent of students in the United States receive some sort of gifted educational programming. School must think about how giftedness is defined and measured in order to make decisions regarding which students receive such programming (Pfeiffer, 2015). Ability testing remains one of the most common identification practices in schools (McClain & Pfeiffer, 2012) . This is because ability tests are thought to be a measure of academic potential (Chapmann, 1988). Consistent relationships exist between individually-administered ability tests and academic achievement (Caemmerer, Maddocks, Keith, & Reynolds, 2018; Deary, Strand, Smith & Fernandes, 2007; Kaufman, Reynolds, Liu, Kaufman & McGrew, 2012). Group-administered ability tests provide an efficient, cost-effective, alternative to traditional intelligence tests when assessing students for gifted programming (Cao, Jung, & Lee, 2017). However, research associating group-ability tests and academic achievement tests is limited, and their ability to measure students' academic potential has not fully been explored. This study targeted this gap in literature by exploring ability–achievement relationships between two commonly administered group-ability tests, the Cognitive Abilities Test (CogAT) (Lohman, 2012) and the Naglieri Nonverbal Ability Test (NNAT) (Naglieri, 2018), and individual academic achievement in Reading and Mathematics as measured by performance on the Weschler Individual Achievement Test, Third Edition (WIAT-III) (Weschler, 2009). This study in Seventy-five second-eighth grade students who had received group-ability testing through their local schools or a university training clinic received follow-up educational testing with the WIAT-III. Using regression analysis, this study found that both the NNAT and the CogAT relate to academic achievement in reading and mathematics. Furthermore, the three distinct batteries of the CogAT (Verbal, Quantitative, and Nonverbal) were found to differentially relate to reading and mathematics achievement. The findings from this study provide support for the use of these group-administered ability tests as a measure of academic potential which may be useful in the identification of students eligible for gifted programming. Schools should consider these findings, amongst other student characteristics, when making educational placement decisions based upon students' performance on these group-administered assessments.

CHAPTER 1: CONCEPTUALIZATIONS OF GIFTEDNESS

Any discussion of what it means to be *gifted* must begin by acknowledging that giftedness is a social construction – a "created concept that is useful and can be operationally defined and measured" (Pfeiffer, 2015, p. 7). Broadly speaking, conceptualizations of giftedness fall within three categories: psychometric models, systems models, and developmental models (Sternberg & Kaufman, 2018). Differences between these categories of conceptualizations include "the importance of non-intellectual abilities, the role of creativity in giftedness, and whether giftedness is potential or achievement" (Sternberg & Kaufman, 2018, p. 37). These three models of giftedness will be discussed below.

Psychometric Models of Giftedness

Models of giftedness that emphasize the measurement of intelligence and consider high intelligence as the hallmark of giftedness are referred to as psychometric models. The earliest notions of what it means to be gifted are rooted in the measurement of intelligence. In the early 1900's Charles Spearman observed that different cognitive tests correlated with each other (Spearman, 1904). Using factor analysis, Spearman labeled the common variance across all cognitive tests "*g,"* or general intelligence. Around the same time that *g* was "discovered," schools worldwide were looking for ways to measure children's cognitive abilities in order to make educational placement decisions. Alfred

Binet and Theodore Simon developed a mental scale for the purpose of identifying children who should receive alternative education in France (Binet & Simon, 1916).

Lewis Terman adapted the work of Binet and Simon and created the Stanford Binet Intelligence Scale (SBIS; Terman, 1916), which became the first widely used intelligence test in the United States. Similar to Binet and Simon's (1916) work in France, this test was utilized to make educational placement decisions as American schools underwent reform. Administration of the SBIS produced only one score – the intelligence quotient (IQ). This score quantified the *general* cognitive ability of an individual. The interpretation of the intelligent quotient classified individuals into descriptive categories of ability. These descriptive categories were age-based comparisons of ability. Terman's classification scheme labeled individuals with an IQ of above 135 as "moderately gifted" and individuals with IQ's above 150 as "exceptionally gifted" (Terman, 1925). Thus, the concept of giftedness arose from the measurement of intelligence.

Although early measurement focused on *general intelligence*, not all psychologists believed that a general factor best represented the construct of intelligence. For example, Thurstone (1938), using a different method of factor analysis, identified seven statistically independent primary mental abilities. The debate between theorists such as Thurstone and Spearman was not resolved from a theoretical standpoint, but amounting statistical support resulted in hierarchal models of intelligence (with a general ability at the top, and specific facets of intelligence below it) becoming the most dominant way to model cognitive abilities (Flanagan & Harrison, 2012). As a result, two models of intelligence began to dominate intelligence theory.

First, Cattell proposed that the construct of intelligence was best represented by two, not one, general factors: fluid intelligence and crystallized intelligence (1943). This became known as the Gf-Gc theory of intelligence. Cattell's theory was influenced by Hebb's (1942) work which differentiated between innate intelligence and intelligence acquired through educational experiences. Cattell (1943) defined fluid intelligence as reasoning abilities to solve novel problems, whereas crystallized intelligence is the storage and retrieval of prior knowledge. In 1963, Cattell experimentally provided support for his theory. Analysis of children's performance across nine cognitive tasks revealed that each task loaded on either a Fluid or Crystallized intelligence factor. Although his analysis revealed that a third-order factor was appropriate, Cattell chose not include a general factor in his model. Cattell's (1963) findings were replicated by Horn (1965), Cattell's student, who identified several other second-order factors in addition to Fluid and Crystalized Intelligence. From this point on, Cattell and Horn collaborated and expanded upon the Gf-Gc model. Today, the model includes nine broad abilities: fluid intelligence, crystallized intelligence, short-term acquisition, visual intelligence, auditory intelligence, long-term storage and retrieval, cognitive processing speed, correct decision speed, and quantitative knowledge.

The second dominate view of intelligence is Carroll's (1993) Three Stratum Theory of Intelligence. In his publication, *Human Cognitive Abilities: A Survey of Factor–Analytic Studies* (Carroll, 1993) concluded that a three-tiered model best represented the construct of intelligence. This model became known as the Three Stratum

Theory of Intelligence. In this theory, general intelligence (*g*) is the only ability placed at the third level, highlighting that all other cognitive abilities and skills fall below it (Carroll, 1993). Just beneath *g*, at the second level, lie eight "broad abilities": fluid intelligence (G*f*), crystallized intelligence (G*c*), general memory and learning (G*y*), broad visual perception (G*v*), broad auditory perception (G*a*), broad retrieval ability (G*r*), broad cognitive speediness (G*s*) and reaction time and decision speed (G*t*) (Carroll, 1993). Beneath these abilities, at the lowest level, are more than 50 "narrow cognitive abilities" which are organized beneath their corresponding broad abilities (Carroll, 1993).

The Gf-Gc model and the Three-Stratum Theory have since been consolidated into the Cattell–Horn–Carroll (CHC) model of intelligence (Schneider & McGrew, 2012). In CHC theory, general intelligence (*g*) lies at the highest level, just as it does in Carroll's (1993) Three Stratum Theory. However, according to CHC theory, the concept of *g* is not particularly meaningful in applied clinical contexts (Schneider & McGrew, 2012). Broad abilities, which fall at the second stratum or level, are considered to be much more applicable to assessment and applied clinical settings according to CHC theory. Interestingly, in Carroll's original writings, he never dismissed the importance of *g* in clinical settings (Benson, Beaujean, McGill, & Dombrowski, 2018; McGill, Dombrowski, & Canivez, 2018). CHC theory has evolved from eight broad abilities to 16 broad abilities (Schneider & McGrew, 2012). The most commonly referenced in terms of clinical application are:

- **General Comprehension Knowledge (G*c*)**: "the breadth and depth of knowledge and skills that are valued by one's culture" (Schneider &

McGrew, 2012, p. 122). G*c* is often assessed through tasks that aim to measure students vocabulary acquisition, listening skills, ability to communicate clearly (Schneider & McGrew, 2012).

- **Fluid Reasoning (G*f*):** "the deliberate but flexible control of attention to solve novel 'on-the-spot' problems that cannot be performed by relying exclusively on previously learned habits, schemas and scripts" (Schneider & McGrew, 2012, p. 111). G*f* is often measured in by tasks that require individuals to form concepts, classify unfamiliar items, apply old rules to new problems, and generate hypotheses (Schneider & McGrew, 2012).
- **Visual Spatial (G*v*):** "the ability to make use of simulated mental imagery (often in conjunction with currently perceived images) to solve problems (Schneider & McGrew, 2012, p. 129). G*v* abilities include mentally simulating how complex patterns and shapes might look when transformed, mentally producing vivid images, and visualizing a path or directions (Schneider & McGrew, 2012).
- **Short-Term Memory (G*sm*):** "the ability to encode, maintain, and manipulate information in one's immediate awareness" (Schneider & McGrew, 2012, p. 114). One key aspect of G*sm* is working memory, the ability to hold information and perform relatively simply manipulations while avoiding distractions (Schneider & McGrew, 2012).
- **Processing Speed (G*s*):** "the ability to perform simple, repetitive cognitive tasks quickly and fluently" (Schneider, McGrew, 2012, p. 119).

It is the measurement of the speed at which individuals complete a task that takes little mental effort, such as discriminating between visual stimuli.

- **Long-Term Storage and Retrieval (G*lr*):** "the ability to store, consolidate, and retrieve information over periods of time" (Schneider & McGrew, 2012, p. 116). G*lr* can be separated into Learning Efficiency and Retrieval Fluency. Measures of Learning Efficiency often assess individuals associative memory, meaningful memory, or free recall of newly learned information (pairs of words, story information). Retrieval fluency tasks require students to retrieve information from long-term memory (Schneider & McGrew, 2012).

CHC theory has profoundly impacted the way the construct of intelligence is measured. Today, many current editions of standardized intelligence tests are heavily influenced by CHC theory, and the measurement of broad abilities (Keith & Reynolds, 2010). For example, administration of the Wechsler Intelligence Scale for Children, Fifth Edition (Wechsler, 2013) yields index scores that correspond to six of the stratum II CHC broad-abilities. The Woodcock-Johnson's Cognitive Battery's technical manual explicitly states that the test was developed around CHC theory, and provides interpretive suggestions in-line with CHC research (McGrew, LaForte, & Schrank, 2014).

It is important to note, that while CHC is not a conceptualization of giftedness in itself, it has played an important role in the conceptualization of giftedness by "suggesting that beneath *g*, there are hierarchically related abilities that contribute to intellectual gifts" (Sternberg & Kaufman, 2018, p. 32). This suggests that individuals can be conceptualized as gifted even if their gifts are within the measurement of broad cognitive abilities rather than general intelligence, as measured by standardized assessments.

CHC's emphasis on broad abilities has sparked debate on which abilities are most relevant to the construct of giftedness. Namely, the debate regarding the importance of nonverbal and verbal intelligence in the conceptualization of giftedness. Nonverbal intelligence is defined as the ability to analyze information and solve problems without language. Whereas verbal intelligence utilizes language-based reasoning. Verbal intelligence is assumed to be transmitted through culture, whereas nonverbal intelligence is often assumed to be more independent of cultural influences. This distinction is rooted in the G*f*–G*c* model of intelligence described earlier, in which one factor of intelligence is unlearned (G*f*) and the other is considered to be learned (G*c*). Further, the ideas of verbal and nonverbal intelligence are embodied in CHC theory, where some broad abilities are, by definition, impacted by culture (general comprehension knowledge). Whereas others, namely fluid reasoning, are thought to be more "culture-fair" (Giessman, Gambrell & Stebbins, 2013). The "culture-fair" label is a step-back from the previous tendency to label tests as "culture-free" (Lohman 2005). As modern theorists acknowledge it is impossible to eliminate the effects of culture when measuring intelligence (Lohman, 2005). It is important to note that many of the key figures in the early conceptualizations, such as Terman, believed that the construct of intelligence was innate and rooted in biology (Jolly, 2018). Thus, early conceptualizations of giftedness placed emphasis on cognitive abilities that are conceived to be less influenced by culture and language. However, more modern conceptualizations of giftedness have begun to embrace more culture-dependent abilities when defining giftedness.

Systems Models of Giftedness

Those who hold purely psychometric views of giftedness tend to de-emphasize the role of psychological processes (i.e. motivation, creativity). According to VanTassel-Baska (2005) this is because those who operate from the psychometric view believe that psychological processes, such as creativity, are the product of high intelligence, rather than a characteristic that interacts with high intelligence. More recent conceptualizations of giftedness acknowledge the role of psychological processes, and view giftedness as a system in which psychological variables interact with cognitive abilities to produce giftedness.

WICS Model of Giftedness

One conception of giftedness that falls within this categorization is Sternberg's (2003; 2005) WICS model. In this model giftedness is conceptualized as the interaction between wisdom, intelligence, and creativity (WICS). Sternberg (2003; 2005) theorizes that creativity is necessary for producing novel ideas. Individuals must also have academic and practical intelligence in order to assess and promote these ideas (Sternberg, 2003; 2005). The construct of wisdom is often criticized for being unable be neatly operationalized. Sternberg and Kaufman (2018) summarize wisdom as the "use of one's abilities and knowledge to achieve a common good" (p. 34). This means that individuals with wisdom are able to harness their abilities and knowledge to achieve something that benefits society, rather than just themselves. One criticism of the WICS model is that it offers no framework for how to identify, or measure, giftedness in children (Sternberg & Kaufman, 2018). For this reason, other systems theories have gained more popularity in practice.

The Three Ring Conception of Giftedness

One particularly popular theory is Renzulli's Three Ring Conception (1978;2005). Renzulli (2005) asserts that there are two types of giftedness – schoolhouse giftedness and creative productive giftedness. Schoolhouse giftedness refers to high test scores, and is compatible with the psychometric conceptualizations of giftedness previously discussed (Pfeiffer, 2015; Renzulli, 2005). Creative productive giftedness can be seen as the type of giftedness that produces gifted behaviors, and the production of "creative and extraordinary things in culturally valued fields" (Pfeiffer, 2015, p. 29). Both types of giftedness are encompassed in Renzulli's Three-Ring conceptualization (2005), which views giftedness at the interaction of three characteristics: above-average ability, task-commitment and creativity. Renzulli (1978) defined task-commitment as a focused motivation. Whereas motivation is often defined by psychologists as a general energy that results in general behaviors, the motivation embodied in the term task commitment is energy towards a specific task.

Renzulli's (2005) more liberal approach to intelligence's role in giftedness is the result of research that supports a "threshold effect." The threshold effect states that beyond certain levels of intelligence, degrees of accomplishment become only weakly associated with intelligence (Chambers, 1969; Stein, 1968; Walberg, 1969;1971).

According to Renzulli (2005), once an individual's IQ is above 120, other (non-cognitive) variables become increasingly important. It is worth reiterating just how liberal this is in comparison to traditional psychometric views of giftedness. That is, Renzulli's threshold of 120 is one standard deviation lower than Terman's (1925) classification of "moderately gifted". This means that almost ten percent more children possess the level of cognitive ability necessary to be considered gifted in comparison to Terman's original classification scheme.

Developmental Models of Giftedness

Developmental models of giftedness posit that giftedness is a dynamic construct, influenced by both internal and external factors, that develop over time (Sternberg & Kaufman, 2018). The inclusion of external factors set developmental models apart from psychometric and systems models of giftedness. For example, Monks (1992) added external factors such as school, family, and peers to Renzulli's (1978; 2005) Three Ring Conception (above average ability, task-commitment and creativity) in his Multifactor Model of Giftedness. Other conceptualizations of this category include the work of Tannenbaum (1986), Feldhusen (1998), Feldman (2000), and Gagné (2005). While there are differences between the work of each theorist, all theorize that individuals have some pre-determined abilities (intelligence) that interact with psychological factors and environmental factors to produce giftedness. The way that these abilities and factors interact is not conceived as stable throughout an individual's development. In other words, these interactions change over time.

Developmental models also deviate from psychometric and systems models by emphasizing accomplishments rather than potential. Subotnik (2009) asserts that developmental models of giftedness shift away from placing emphasis on *potential* in childhood, towards placing emphasis on *accomplishment* in adolescence and adulthood. This is not evident in other models of giftedness, which consider giftedness as a stable trait. For example, under the purely psychometric perspective an adult with a very high IQ is recognized as gifted, regardless of their accomplishments (Pfeiffer, 2015). Subotnik (2009) would argue that accomplishment is necessary to be labeled gifted as an adult. This idea is embodied in both Gagné's (2005) and Tannenbaum's (1986) work. Gagné (2005) highlights the distinction between potential and achievement in his Differentiated Model of Gifted and Talented Development (DMGT) in which he distinguishes between "gifts" (genetically-determined) and "talents" (domain-specific accomplishments). Gagné's (2005) model strived to identify how environmental influences turn gifts into talents. Similarly Tannenbaum (1986) conceptualizes giftedness as something an individual develops into (rather than is).

CHAPTER 2: GIFTED EDUCATION AND IDENTIFICATION

Gifted education in the United States can be traced back to the late 19th and early 20th centuries (Jolly, 2018). This time, referred to as the Progressive Era in American history, was a period of widespread social activism and political reform. Educational reform was one of the prime concerns for activists of this era. As a result, the American education system experienced a period of rapid growth, which is often referred to as the Progressive Era of Education (Jolly, 2018). This growth was the result of growing societal concern regarding the use of child-labor, and the lack of education for working-class youth (Jolly, 2018). During this time, extensive child labor laws, as well as compulsory education laws were enacted; limiting the child work force, and increasing school enrollment. Furthermore, immigration during this time period also contributed to a larger student population (Jolly, 2018). This rapid growth resulted in many challenges. Namely, the task of educating a student body that was larger, and more diverse than ever before (Jolly, 2018).

During the Progressive Era, the field of psychology as a science also emerged in the United States. Psychologists found themselves eager to contribute, and provide solutions for the new challenges faced by educational systems. This gave rise to several branches of psychology including Educational and School Psychology, and eventually led to the sub-field of gifted education. Gifted education is concerned with "effecting desirable changes in our most able students through proper educational provisions and adaptations" (Dai & Chen, 2013, p. 152)

The sub-field of gifted education is not unified. In fact, researchers note a lack of conceptual agreement and a disconnect between theory and research (Ziegler & Raul, 2000; Ambrose, VanTassel-Baska, Coleman & Cross, 2010). Dai and Chen (2013) have categorized approaches to gifted education into three paradigms: the gifted child paradigm, the talent-development paradigm and the differentiation paradigm. Each paradigm frames gifted education as a set of assumptions, goals, and proposes identification and educational practices (Dai & Chen, 2013). Assumptions in each paradigm are based off of the different conceptualizations of giftedness. Goals in each paradigm outline each paradigms' purpose for gifted education. Identification refers to each paradigms criteria for identifying students eligible for gifted education. These criteria are consistent with the assumptions and goals of each paradigm. Lastly, educational practices reflect the how goals of each paradigm will be met through instruction.

The Gifted Child Paradigm for Gifted Education

Under the Gifted Child Paradigm, it is assumed that giftedness is an individual characteristic that can be identified through intelligence testing (Dai & Chen, 2013). This assumption pays allegiance to psychometric conceptualizations of giftedness that view high intelligence as the hallmark of giftedness. In the United States this paradigm is largely a product of the simultaneous education reform and development of intelligence tests in the early 1900's (Jolly, 2018). During this time, intelligence tests were conceived as a way to measure student potential, and therefore were thought to be able to sort children into appropriate educational environments (Chapman, 1988). Under the gifted education paradigm, the characteristic of high intelligence makes gifted students qualitatively different from their non-gifted peers and they are assumed to be able to learn academic content faster and be able to master more complex content. In other words, children who rank high on intelligence tests are thought to have greater academic potential. Under the gifted child paradigm the goal of gifted education is to harness this potential (Dai & Chen, 2013).

Unsurprisingly, identification practices under this paradigm are predominately the use of various ability (intelligence) tests to determine which children possess the characteristic of being gifted. Consistent with the psychometric view, the gifted child paradigm posits that only children whose performance ranks at the high-end of the normal distribution of these tests are to be considered “gifted”.

Traditional ability tests are standardized assessments that are administered to individual students by trained professionals (Cao, Jung, & Lee, 2017). Cao, Jung and Lee (2017) highlight that administering individually-administered assessments when identifying students for gifted program can be advantageous due to their one-on-one nature. For example, an examiner is able to make behavioral observations regarding a student’s attention and problem-solving strategies (Cao, Jung, & Lee, 2017).

However, the administration of traditional intelligence tests can be lengthy. For this reason, abbreviated ability tests are also used in identifying students for gifted educational programming (Cao, Jung, & Lee, 2017). These shorter-versions raise some concerns. Namely, research has shown that these abbreviated batteries can produce very different scores than their longer counterparts (Pierson et al., 2012). Therefore, several researchers caution the use of such measures, despite their efficiency (Kruyen, Emons & Sijtsma, 2013; Smith, Combs, & Pearson, 2012).

America's entry into World War I influenced identification practices (Jolly, 2018). During World War I, group-administered ability tests were utilized in military recruitment efforts with much success. America's entry into World War I influenced identification practices (Jolly, 2018). During World War I, group-administered ability tests were utilized in military recruitment efforts with much success. This led to an increase in the use of group-administered ability tests in education (Jolly, 2018). Group ability tests can be administered simultaneously to large groups of students and are able to be administered by administrators who are not psychologists. Group-administered tests are considered to be a more efficient, practical, and cost-effective way to identify children for gifted education in comparison to traditional ability tests (Cao, Jung, & Lee, 2017).

A distinction is also made between verbal and nonverbal ability tests. Nonverbal ability tests are tests whose stimuli are presented visually through the use of concrete objects, line drawings, or spatial visualization. Further, nonverbal tests do not require verbal responses. However, this is not to say that these tests don't require verbal processes. As Lohman (2015) states nonverbal tests often "use tasks whose solution is greatly facilitated by the use of verbal or mathematical cognitive processes" (p. 113). Commonly administered individual nonverbal assessments include: Weschler Nonverbal

Scale of Ability (WNV; Weschler & Naglieri, 2006), the Universal Nonverbal Intelligence Test (UNIT; Bracken & McCallum, 1998), the Test of Nonverbal Intelligence (TONI; Brown, Sherbenou, &Johnsen, 2010). Commonly administered group-nonverbal assessments include the Naglieri Nonverbal Ability Test (NNAT; Naglieri, 2018) and the Nonverbal Battery of the Cognitive Abilities Test (CogAT; Lohman, 2012). Nonverbal intelligence tests have become increasingly popular in the identification of gifted children, as some believe they may lead to more equitable identification of culturally and linguistically diverse students, especially when used as universal screeners (Giessman, Gambrell & Stebbins, 2013; Naglieri & Ford, 2015). However, research appears to consistently find group-differences between minority and majority populations in nonverbal measures (Giessman, Gambrell & Stebbins, 2013; Lohman, Korb & Lakin, 2008).

Today, intelligence tests (individually or group-administered) appear to be the most common way to identify students for gifted programming in the United States (McClain & Pfeiffer, 2012). This is evidenced by the fact that the use of a global IQ score remains a dominant criterion used for gifted education (Feldhausan & Jarwan; 2000; McClain & Pfeiffer, 2012). McClain and Pfeiffer (2012) found that 18 states require specific test scores for students to qualify for gifted programs. All 18 of which require specific cut-off scores on intelligence tests. The most frequent cut-off scores used in identification place a child's performance within the top 3-5 percent, consistent with early psychometric models of giftedness (Borland, 2005; Pfeiffer, 2015). These identification practices suggest that many states, to some extent, operate from a gifted child paradigm when providing gifted services to their students.

Because so much of the gifted children paradigm is concerned with assuming that gifted students are qualitatively different from their non-gifted peers, gifted education is conceived as a qualitatively different education – with a different, usually accelerated curriculum (Dai & Chen, 2013). Rodgers, (2007) noted the use of subject-based and/or grade-based acceleration as the main approach to educating children who are assumed to learn at a faster pace.

The Talent Development Paradigm for Gifted Education.

While intelligence tests were readily used as a solution to the challenges faced by education systems during the progressive era, disagreements about their use existed and discontent with purely psychometric views of giftedness steadily grew. As a result, views of gifted education gradually expanded beyond the gifted child paradigm. The talent-development paradigm does not exclude the role of intelligence, but rather assumes a "broader psychosocial" (p. 156) basis of potential (Dai & Chen, 2013). Because of this, the talent-development paradigm is rooted in systems conceptualizations of giftedness, which place emphasis on psychosocial characteristics such as creativity and motivation.

The expansion of gifted education away from purely psychometric views was cemented in the Marland Report of 1972. The Marland Report (1972) established the first federal definition of gifted students as children who have "demonstrated achievement and/or potential ability" (p. 10) in one or more of the following domains: general intellectual ability, specific academic aptitude, creative or productive thinking, leadership, visual or performing arts, and psychomotor ability (Marland, 1972). As a result, the talent-development paradigm of gifted education steadily grew.

It appears as though school-systems are increasingly embracing this paradigm. Although the use of intelligence and academic tests are certainly dominant McClain & Pfeiffer's (2012) also concluded a dramatic shift towards views consistent with the talent development paradigm for gifted education. Between the years of 2000 and 2010, considerable change occurred the terminology that states and districts used to define what it means to be gifted. Specifically, states and districts have increased using the term "creativity" (McClain & Pfeiffer, 2012). Today, most states use a multi-dimensional approach when identifying students; signaling that educators conceive giftedness as being composed of more than high intelligence (McClain & Pfeiffer, 2012).

The Differentiation Paradigm for Gifted Education

The assumption that underlies the differentiation paradigm differs from the other paradigms in that it is now an assumption directly related to conceptualizations of giftedness. Instead, the differentiation paradigm assumes that a students' curricular content should be within their zone of proximal development (Dai & Chen, 2013). By extension, it is assumed that the general curriculum is at the lower limit of some (gifted) students proximal development, or in other words too easy. The purpose of education in the differentiation paradigm is to provide students with instruction that is appropriately challenging (within their zone of proximal development). In the differentiation paradigm, gifted education is achieved by providing a "dynamically response educational match" (Dai & Chen, 2013, p. 158) to high-achieving students that are left unchallenged by the general curriculum.

The differentiation paradigm has emerged as the result of the changing landscape of service delivery in special education. Namely, the development of the Response to Intervention (RtI) framework in service delivery. At its core, RtI is a tiered service delivery model aimed to prevent and remediate academic difficulties through effective classroom and supplemental instruction (Buysse & Peisner-Feinberg, 2013). Many different models of RtI exist, but most models use three tiers. The three-tiered approach is modeled after public health models that use primary, secondary, and tertiary prevention (Buysse & Peisner-Feinberg, 2013). The application of a tiered model to education refers to the different types of instruction that are used with students (Brown-Chidsey et al., 2009). At the base of an RTI model (Tier 1) is the general curriculum that all students receive. At each tier, the nature of the instruction/intervention becomes more intense (Brown-Chidsey et al., 2009; Buysee & Peisner-Feinburg, 2013; Fuchs & Fuchs, 2006). According to Fuchs and Fuchs (2006), the increase in intensity across tiers is achieved in five ways: (1) using more teacher-centered, systematic, and explicit instruction; (2) conducting it more frequently; (3) adding to its duration; (4) creating smaller homogenous groups; or (5) relying on instructors with greater expertise.

Some researchers have explicitly extended the RtI framework to gifted education. Bianco (2010) adapted RtI's three-tier model to include the assessment of gifted students. In Bianco's (2010) writings, students in tier 1 are provided instruction that meets the needs of average students in a general mixed-ability class. Students in the second tier are provided with differentiated instruction that provides a challenge for students who are unchallenged by tier 1 instruction (Bianco, 2010). In Tier 3, the most capable students are provided with intense interventions such as advanced classes (2010). Choice (2011) extended the RtI triangle in a similar manner, she calls "Response to Intelligence."

Identification practices under the differentiation paradigm identify the levels of need within in students. This is different than other paradigms in which goal of identification practices identify a group of students as "gifted." Identification practices under the differentiation paradigm can be achieved in many ways, such as high-ceiling tests and curriculum-based measures (Matthews & Foster, 2006; Reis, Burns & Renzulli, 1992). McClain and Pfeiffer (2012) found that no states had identification practices that involved reoccurring assessments of giftedness. Therefore, it appears that the school systems are not actively operating from the differentiation paradigm of gifted education.

CHAPTER 3: INTELLIGENCE AND ACADEMIC ACHIEVEMENT

As previously discussed, gifted education and the measurement of intelligence in the United States arose at the same time (Jolly, 2018). Intelligence tests were conceived to be able to sort children based on their assumed ability to measure an individual's *potential* to achieve (Chapmann, 1988). Indeed, relationships have been consistently found between general intelligence and academic achievement in the research literature (Caemmerer, Maddocks, Keith, & Reynolds, 2018; Deary, Strand, Smith & Fernandes, 2007; Kaufman, Reynolds, Liu, Kaufman & McGrew, 2012).

Recently, researchers have focused on ability–achievement relationships through the lens of CHC theory, which focuses more on broad abilities rather than *g*. In review of the CHC-Achievement research, McGrew and Wendling (2010) were concluded that broad "cognitive abilities contribute to academic achievement in different proportions in different academic domains, and these proportions change over the course of development" (Schneider & McGrew, 2012, p. 108).

Reading

According to McGrew & Wendling (2010) several broad CHC cognitive abilities predicted basic reading skills at one or more age group. Basic reading skills include decoding and word recognition (McGrew & Wendling, 2010).

General comprehension knowledge (G*c*) has been found to moderately predict basic reading in childhood. McGrew and Wendling found that the G*c*–basic reading link could be attributed to specific narrow abilities under G*c*. The narrow ability of general verbal information(G*c*-K0) was consistently related to basic reading skills, and it's importance has been found to increase with age (McGrew & Wendling, 2010). This finding is consistent with research that highlights prior background knowledge's role in reading development (Kintsch & Rawson, 2005). Furthermore, the narrow ability of Listening Ability (G*c*-LS) was found to be moderately related to basic reading skills in early elementary aged children (McGrew & Wendling, 2010). This finding supports previous research that notes the importance of listening comprehension in reading development (Hoover & Gough, 1990; Joshi & Aaron, 2000).

Long-term storage and retrieval (G*lr*) is also linked to basic reading skills (McGrew & Wendling, 2012). Specifically, being able to retrieve letter–sound relations is crucial for basic reading tasks (Vellutino, Tunmer, Jaccard & Chen, 2007). The relationship between G*lr* and basic reading declines with age. This is hypothesized to occur because as students as students' progress through learning how to read, the retrieval of letter–sound parings becomes automatic.

Similarly, relationships between Processing Speed (G*s*) and basic reading appear to decline with age (McGrew & Wendling, 2010). Specifically, performance on naming speed or any other processing speed related measures have been linked to early reading skills. For example, rapid automatic naming – the speed at which children can quickly and accurately name an array of common visual stimuli – is associated with basic reading skills (Araujo, Reis, Petersson & Faisca, 2014).

Another aspect of reading is reading comprehension, which can be defined as the ability to construct meaning from text. Again, McGrew and Wendling (2010) identified several broad CHC cognitive abilities (auditory processing, comprehension knowledge, long-term retrieval, and short-term memory) predicted reading comprehension at one or more age-group.

Comprehension Knowledge (G*c*) abilities have been established as a strong predictor of reading comprehension (McGrew & Wendling, 2010). This relationships is consistent at different ages (McGrew & Wendling, 2010). An extensive body of research suggests that general language development, vocabulary, and prior knowledge, which are all of which are measured when assessing G*c*, are important to students' reading comprehension abilities (Jenkins, Fuchs, Van den Broek, Epsin & Deno, 2003, Kintsch & Rawson, 2005; Nation, 2005; Perfetti, 2007). So much so, that Floyd, Bergeron and Alfonso (2006) stated that the biggest difference between individuals with good and poor reading comprehension are measured G*c* abilities.

Short term memory (G*sm*), specifically the narrow ability of memory span (G*sm*-MS) has been found to relate to students' reading comprehension abilities. McGrew & Wendling (2010) hypothesize that this relationship may actually be rooted in the narrow G*c* ability of listening ability (G*c*-LS). This is because many measures of memory span that have been linked to reading comprehension require students to remember lengthy items, such as long sentences. The memory for sentences increases the demand of G*c*-LS abilities in comparison to other rote memory tests (McGrew & Wendling, 2010). This hypothesis reiterates the importance of G*c* in reading comprehension.

McGrew and Wendling classified Fluid Reasoning (G*f*) to be "tentatively" related to reading comprehension abilities in students between the ages of 14 and 19. This classification is the result of inconsistent findings across research studies investigating G*f*–reading comprehension relationships. Studies that show significant relationships between G*f* and reading comprehension highlight that G*f* may only be associated at higher levels or reading comprehension (Floyd et al., 2006; McGrew, 1993; Nation, Clarke, & Snowling, 2002).

McGrew and Wendling (2010) did not find any significant relationships between visual processing (G*v*) and any domain of reading achievement.

Mathematics

Basic math skills include arithmetic and computational skills. They are often assessed through tasks that assess a child's basic math facts and ability to solve complex algorithmic computations. The broad CHC cognitive abilities of comprehension knowledge, fluid reasoning, and processing speed predict achievement in basic math skills (McGrew & Wendling, 2010).

Individually-tested General Comprehension Knowledge abilities were found to be moderately related to basic math skills across school-aged children and adolescents (McGrew & Wendling, 2010). Language skills have consistently been found to play an important role in math skills. Specifically, language skills contribute to the development of number concepts and the ability to retrieve basic math facts (Gelman & Butterworth, 2005; Chong & Siegel, 2008).

Fluid Reasoning (G*f*) is consistently found to be influential to basic math skills (McGrew & Wendling, 2010). Floyd and colleagues (2003) highlight possible explanations for this consistent relationship. One explanation is that both Three Stratum Theory of Intelligence (Carroll, 1993) and CHC theory (Schneider & McGrew, 2012) consider quantitative reasoning to be a narrow ability of G*f*. Another explanation is that G*f* abilities have been found to impact specific math skills, such as rational number calculations, whole-number line estimation and algebra (Namkung & Fuchs, 2016; Seethaler, Fuchs, Star & Bryant, 2011; Singler & Bunge, 2014).

Similarly, Processing Speed's (G*s*) relationship with basic math skills has also been explained by its impact on specific math skills. Processing Speed is hypothesized to determine how quickly students can count numbers (Fuchs et al., 2006). Slower processing speed is thought to impede basic math skills by resulting in slower counting speed (Geary, 2007), slower numerical processing fluency (Swanson & Jerman, 2006), and difficulty with quick and accurate execution of simple math-related cognitive skills (Fuchs et al., 2006; Geary, 2007).

Another aspect of mathematic achievement is often labeled " math problem solving" or "math reasoning." Math problem solving is typically assessed through items like word problems, number series, and problems that require the application of mathematic operations on concepts (McGrew & Wendling, 2010). McGrew and

Wendling's (2010) findings suggest that several broad CHC abilities were significantly predictive of math problem solving abilities at different age groups.

First, Fluid Reasoning (G*f*) was found to highly predict math problem solving between in elementary to middle school-aged students. G*f* continues to significantly predict math problem solving in high-school aged students, however relationship is not as strong at the G*f*-math problem solving link at younger ages (McGrew & Wendling, 2010). Research has consistently indicated that Gf cognitive processes such as concept formation play a role in math problem solving. For example, Fuchs and colleagues (2006) found that performance on the WJ-III's (Woodcock et al., 2001) concept formation task correlated with math problem solving abilities. Nonverbal abilities (as assessed via matrix problems) have consistently been linked to math problem solving ability (Fuchs et al., 2006). Matrix problems are widely considered to be a measure of G*f*.

Second, Comprehension Knowledge (G*c*) was found to increasingly predict math problem solving abilities with age (McGrew & Wendling, 2010). Findings that link G*c* to math problem solving highlight the importance of language abilities. For example, Proctor, Floyd and Shaver (2005) found students identified as having a math disability often have weaknesses in their oral language abilities. Further, math problem solving, often assessed through word problems, requires that students interpret linguistic information to a construct a problem model (Fuchs et al., 2006). As students become older the linguistic demands of items used to measure math reasoning increase, and thus G*c* abilities become more importance (Fuchs et al., 2006).

McGrew and Wendling (2010) did not find any significant relationships between visual processing (G*v*) and mathematics achievement. This is commensurate with findings from Flanagan et al. (2006), which also found no significant G*v* relationships. This is a bit surprising given that G*v* abilities have been found to be an area of weakness for individuals identified as having math disabilities (Osmon, Smertz, Braun, & Plambeck, 2006). Several answers to this "G*v* mystery" (McGrew & Wendling, 2010, p. 665) have been proposed. McGrew & Wendling suggest that this lack of a relationship may be explained due to specification error (the way we measure visual processing abilities does not measure aspects of G*v* related to achievement). On the other hand, Flanagan et al. (2006) suggest that G*v* abilities may only be important for specific higher level math, such as geometry and calculus.

CHAPTER 4: GROUP-ABILITY TESTS AND CHC THEORY

The Cattell–Horn–Carroll (CHC) theory of intelligence has not only influenced how researchers investigate ability–achievement relationships, but it also heavily influenced the way ability is measured. CHC theory has become the foundation of many recently developed tests of cognitive abilities. Keith and Reynolds (2010) noted that even assessments without explicit reference to the theory pay some sort of "allegiance" to CHC (p. 635). This influence not only applies to individual assessments of ability, but also carries into how group–ability tests are designed. Both the Naglieri Nonverbal Ability Test, Third Edition (NNAT-3) and the Cognitive Abilities Test, Form 7 (CogAT), two of the most popular group-administered ability tests, both reference aspects of CHC in their manuals (Naglieri, 2018; Lohman, 2012).

In the NNAT-3's manual, Naglieri (2018) states that the items on the NNAT-3 require a type of thinking that involves "seeing relationships among components of the question and thinking of rules that can explain those relationships" (p.2). Further, Naglieri states that this kind of thinking – "in modern conceptions of ability" (p. 2) – is "closely related to general ability". Therefore, while not explicitly mentioning CHC theory, it can be inferred that NNAT-3 can be conceptualized as measuring the broad ability of Fluid Reasoning (G*f*). Naglieri's (2018) claim that this is related to general ability is valid. In fact, Fluid reasoning (G*f*) and general ability (*g*) have sometimes been reported to be perfectly correlated (Floyd, Evans, & McGrew, 2003; Gustafsson, 1984). This has led to some researchers to believe they are the same construct. Horn & Blankson (2012) state that from a theoretical standpoint, G*f* is the same as Spearman's conceptualization of *g*.

The CogAT's Research and Development Guide (Lohman, 2012) explicitly states the influence CHC theory had on the structure of the test. Similar to Naglieri (2018), Lohman (2012) states that the broad ability of Fluid Reasoning (G*f*) is closest to *g* in comparison to other broad abilities. The CogAT's composite score is described as a measure of overall cognitive ability (*g*). Further, Lohman (2012), citing Carroll's (1993) work, asserts that G*f* can be separated into three sub-factors: Verbal, Quantitative and Inductive Reasoning. This is Lohman's rationale for having three distinct batteries of the CogAT (Verbal, Quantitative and Nonverbal). Each battery is said to measure fluid-analytic skills as well as domain-specific abilities.

Lakin and Gambrell (2012) examined the latent factor structure of the forms of the CogAT that utilize picture items using a bifactor MIRT model to separate general (composite) and domain (battery) factors. The results of this study concluded that the picture-based formats were able to isolate between general and domain-specific abilities.

The subtests that make-up the CogAT's verbal battery were found to have the strongest domain factor (Lakin & Gambrell, 2012). Although subtests such as Picture Matrices and Picture Classification require the narrow G*f* ability of induction, they also require a student to possess the narrow ability of general verbal knowledge – "knowledge that one's culture deems essential, practical, and worthwhile for everyone to know"

(Schneider & McGrew, 2012, p. 123). For example, in an item that requires students to choose a ball that is played in a team sport, one must have the background knowledge of sports to be successful. Thus, a child's performance on the verbal battery still, to an extent, depends on their comprehension knowledge (G*c*) abilities. Interestingly, the Sentence Completion was found to have the weakest loading on the CogAT's general factor (Lakin & Gambrell, 2012) . On the forms of the CogAT Lakin & Gambrell (2012) investigated, Sentence Completion is the only subtest to have items that are not completely picture-based. On this task, students have to listen to sentence before choosing the correct picture answer. This task requires a student to use different narrow abilities that fall under broad ability of general comprehension knowledge, such as lexical knowledge and grammatical sensitivity. Sentence Completions strong loading on the Verbal factor, and weak loading on the CogAT's general factor further supports that the Verbal factor of the CogAT requires the use of G*c* abilities, in addition to the G*f* abilities. It may be hypothesized that G*c* becomes more important on levels of the CogAT that do not utilize picture items, and instead use written words.

All of the subtests on the quantitative battery moderately or highly loaded onto a domain factor (Lakin & Gambrell, 2012). All of the subtests on the CogAT's quantitative battery require G*f*. Number Puzzles and Number Series subtests were found to load most strongly on the Quantitative domain factor. Unsurprisingly, these tasks require the narrow G*f* ability of quantitative reasoning. The Number Analogies perhaps relies more on G*f* narrow abilities of induction and deduction, but still requires the application of quantitative concepts.

All of the subtests that make up the CogAT's nonverbal battery loaded onto the general factor, which Lakin & Gambrell conceptualize as general G*f* – but theoretically is similar to *g*. It is not surprising that the Figure Matrices and Figure Classification subtests loaded onto this general factor. Both subtests require the application of narrow the fluid reasoning abilities of induction and general sequential reasoning. It is also not surprising that Paper Folding would load onto a general fluid reasoning factor as research indicates that visualization, a narrow ability of visual-processing defined as "the ability to perceive complex patterns and mentally stimulate how they might look when transformed" (Schneider & McGrew, 2012, p. 111) – load onto both fluid reasoning and visual processing factors (Carroll, 1993). However, the three subtests that make up the nonverbal battery showed weak evidence for a domain-specific factor, with only the Paper Folding subtest loading strongly onto it. Lakin and Gambrell (2012) believe that this may represent a domain factor that is representative of a visual processing (G*v*) abilities. The dual-loading of the Paper Folding Subtest onto both the general and domain factors may be explained by the use of a variety of strategies used to complete the task. Hegarty (2010) states that some people may use mental imagery (i.e. imagining folding the paper, punching the hole, then unfolding the paper) whereas other people may use analytic strategies (using the number of holes/folds to eliminate some answer choices). Schneider and McGrew (2012) believe that G*v* loadings occur when imagery strategies are applied, and G*f* loadings occur when analytic strategies are applied. This may support Lakin and Gambrell (2012)'s conclusion that the domain factor of the nonverbal battery may truly represent narrow visual processing abilities. Although Lakin and Gambrell

(2012) do not study the factor loadings on forms of the CogAT that are meant for older children, the nonverbal battery is composed of the same, but more difficult, tasks on all forms. Therefore, it would be reasonable to believe similar conclusions could be made.

CHAPTER 5: RESEARCH AIMS AND HYPOTHESES

School systems use group–administered cognitive assessments in their screening process for identifying students who may benefit from gifted, or advanced, educational programming (Card & Giuliano, 2015). The Cognitive Abilities Test (CogAT) and the Naglieri Nonverbal Ability Test (NNAT) are two commonly administered group-cognitive assessments utilized for gifted identification purposes (Giessman, Gambrell, & Stebbins, 2013). Despite well-established relationships between individual measures of cognitive ability and academic achievement (Caemmerer, Maddocks, Keith, & Reynolds, 2018; Deary, Strand, Smith & Fernandes, 2007; Kaufman, Reynolds, Liu, Kaufman & McGrew, 2012), little research has been done regarding the relationships between group-administered cognitive assessments and individual academic achievement. The overarching goal of this dissertation is to explore the relationships between cognitive abilities measured by the CogAT and the NNAT and individual academic achievement. This research goal will be accomplished through the following research questions:

1. Are there group-differences in achievement for students who have received “gifted” programming in the form of above-grade level instruction and students who follow a general curriculum?

It is hypothesized that children who have received above-grade level instruction will have higher achievement scores across both Reading and Mathematics achievement.

Furthermore, it is also hypothesized that these differences will be more pronounced in Mathematics Achievement. Namely, Math Calculation skills as math calculation skills are hypothesized to be more responsive to instruction in comparison to more reasoning-based aspects of mathematics. Significant differences will support the use of gifted programming as a control variable.

2. How do students' CogAT Composite scores and NNAT scores relate to reading and mathematics achievement? Are there differences in test–achievement relationships across these different academic domains?

It is hypothesized that the CogAT Composite Score will be a better predictor of both reading and mathematics achievement. This is hypothesized due to the CogAT's broader sampling of cognitive abilities (such as the inclusion of General Comprehension Knowledge (G*c*) in comparison the NNAT (which exclusively measures Fluid Reasoning (G*f*)).

3. How well do students' CogAT Composite Scores relate to academic achievement in comparison to CogAT battery scores (Verbal, Quantitative and Nonverbal)? Which combinations of scores are most relevant to students' academic achievement?

It is hypothesized three distinct batteries of the CogAT (Verbal, Quantitative and Nonverbal) will differentially predict more variance in subject-area achievement in comparison to the CogAT Composite score. This hypothesis is based upon the differential measurement of broad and narrow abilities within each distinct battery, and the known relations between these broad abilities and academic achievement.

Specifically, it is hypothesized that the Verbal Battery of the CogAT will explain the most variance in reading outcomes due to its inclusion of G*c* abilities. It is also hypothesized that the Quantitative Battery of the CogAT will explain the most variance in Mathematics achievement, due to the importance of quantitative concepts in mathematics. Lastly, it is hypothesized that the Nonverbal Battery of the CogAT will predict both reading and mathematics achievement. However, it is predicted to account for less variance in achievement in comparison to the Verbal and Quantitative batteries of the CogAT.

Additionally, hypothesized that in some cases, different combinations of Battery Scores will be the most relevant to academic achievement. It is hypothesized the combination of the Quantitative Battery and the Nonverbal Battery will best predict Mathematics achievement, as this combination embodies both Broad and Narrow Fluid Reasoning (G*f*) abilities.

4. Does the Nonverbal Battery of the CogAT provide significant predictive value above and beyond students' NNAT scores when predicting Reading and Mathematics Achievement?

This research question aims to provide information to schools who wish to use a nonverbal measure when assessing children from diverse cultural and linguistic backgrounds for gifted programming. It has been suggested that the Nonverbal Battery of the CogAT can be used as a stand-alone score when assessing students from culturally and linguistically diverse backgrounds. It is hypothesized that the Nonverbal Battery of the CogAT will not provide significant predictive value above and beyond students'

NNAT scores. This is hypothesized because the CogAT's Nonverbal Battery includes a measure (Paper Folding) of visualization a narrow ability under the broad ability of *Gv* in addition to measures of Fluid Reasoning (G*f*). Current research on ability–achievement relationships does not support G*v* as a predictor of academic achievement.

5. Are the relationships between the batteries of the CogAT, the NNAT and individual academic achievement moderated by the age of the student when achievement testing took place?

It is hypothesized that students' age at the time of achievement testing will moderate the relationships found between group-administered ability tests and academic achievement. Specifically, it is hypothesized that age will significantly moderate the relationship between ability tests scores and Reading Comprehension abilities, so that the relationships are stronger for older children relative to younger children. This is consistent with finding between individually-administered intelligence tests and student reading achievement which find that relationships between intelligence and reading comprehension become stronger with age.

CHAPTER 6: METHOD

Participants

The participants in this study were 75 elementary and middle-school age children. The majority of participants were recruited from a database of children who had received individual or group cognitive evaluations at a university training clinic. The clinic typically conducts roughly 500 cognitive evaluations per year. Most of these evaluations are used towards applications into a local school district's gifted education program. Information about the study was also posted on the university training clinic's website, and therefore not all recruited participants had been previously evaluated at the training clinic.

The participating children were between the ages of 7 and 14 years old. The mean age of the participants was approximately 9.5. Of the 75 participants, 38 (50.7%) were boys and 37 (49.3%) were girls. Parents provided information about their child's race or ethnicity on a background history questionnaire. 44% indicated Caucasian, 33.3% indicated Asian, 12% indicated African American, 0% indicated Hispanic, and 10.6% indicated other. More specifically, the "Other" group consisted of participants who reported to be Mixed-Race (4 participants were reported to be mixed Hispanic/White and 4 participants were reported to be mixed Asian/White). Twenty-seven participants (36.0%) did not receive any gifted programming. Parents also reported if their children qualified for free- or reduced-priced lunch in public schools. Of the 75 participants, only 3(4%) students qualified for free- or reduced-priced lunch. Eighteen participants (24.0%) received part-time gifted programming. 22 participants (29.3%) received full-time gifted programming. The remaining 8 (10.6%) participants were either homeschooled or enrolled in a private school

Procedure

Parents of current second through eighth grade students who had received individual or group cognitive evaluations at a university training center were contacted. Additional parents contacted the research team after reading about the current research study on the training clinic's website. WIAT-III Achievement Testing was scheduled with interested parents. At the appointment, parents signed consent forms and provided the researchers with official score reports from group-ability testing conducted by either their child's local school district or by the university training clinic. Parents also completed a background history questionnaire at this appointment which provided the research team with demographic information, as well as their child's brief educational history. The children also assented to being part of this research study. Children below the age of 8 were read an assent form and asked for verbal assent. Children above the age of 8 were asked to read and sign an assent form independently. Assent took place just prior to WIAT-III achievement testing. Administration of the reading, math, and oral language sections of the WIAT-III took approximately an hour and a half. After this testing appointment, participants' performance on the WIAT-III was written into a report and shared with parents.

Measures

Group-Administered Cognitive Assessments

The Naglieri Nonverbal Ability Test (NNAT)

Participants in this study had previously been administered either the NNAT2 (Naglieri, 2007) or The NNAT3 (Naglieri, 2018) through their local school or the university run training clinic. All students took the most-recent version of the test when they were tested. Both the NNAT 2 and NNAT3 consist of 48 items of increasing difficulty. During test administration, a test proctor reads general testing direction and completes two sample questions with a group of students. Then, students are told to work independently on the remaining test items for a specified amount of time. This administration format is consistent across both versions of the NNAT. There are two formats of items. The missing piece format requires that a student selects the missing piece of a larger image This format is seen on the lower levels of the NNAT that are meant for younger-aged students. The second format involves a series of progressive matrices similar to Figure Matrices subtests of the CogAT (Giessman, Gambrell, & Stebbins, 2013; Naglieri, 2018). Although the NNAT2 and NNAT 3 do not share any of the same questions, the representations of question types is similar across forms of both versions of the test (Naglieri, 2018). Furthermore, performance on the NNAT2 and NNAT 3 was found to be highly correlated (Naglieri, 2018).

The developers of the NNAT3 utilized an alternative-form method of estimating reliability. In the alternate-form method, test developers test the same individuals with different forms of the test that are intended to measure the same thing (Carmines &

Zeller, 2011). In addition to having two-forms, the NNAT3 also has two formats: paper administration and computer-based administration. During test development, two alternate form studies were conducted to estimate the reliability across forms and formats (Naglieri, 2018). In the first study, which analyzed scores from 524 students who took both forms in their paper format within four weeks of each other, correlations between forms averaged *r*=.79 (Naglieri, 2018). The second study differed from the first in that participants took one form of the test in the paper format and one form in the computer-based format (Naglieri, 2018). Using multiple regression moderation analysis it was determined there was a significant mode effect that resulted in higher scores on the paper version by 2-3 points (Naglieri, 2018). This mode effect was incorporated into the conversions from raw-score to scaled-scores to make sure that students of the same ability are calculated to have the same scaled score regardless of which format they take (Naglieri, 2018). Carmine and Zeller (2011) highlight the importance of making such corrections in order to ensure the validity of an assessment.

Test developers of the NNAT3 (Naglieri, 2018) investigated the relationship between the NNAT3 and two other measures of cognitive ability to establish validity. In the first validation study, performance of 368 students on the NNAT3 was compared to their performance of the NNAT2 (Naglieri, 2008). Theoretically, the NNAT3 and NNAT2 measure the same construct, and thus should be highly correlated. This was supported, as correlation coefficients between all levels of the NNAT-3 and NNAT-2 were found to be in the high .70's (Naglieri, 2018). Furthermore, developers on the NNAT-3 also conducted correlation studies between the NNAT3 and the Otis-Lennon

School Ability Test, Eighth Edition (OLSAT 8). Three-hundred and sixty-six students' who took the NNAT3 online as part of the norming sample went on to take the OLSAT 8 within four weeks. Correlations between these two measures revealed that NNAT3 performance was highly correlated to performance on the OLSAT 8 Nonverbal and Total scores (Naglieri, 2018). The NNAT3 had lower correlations with the OLSAT 8's Verbal Score (Naglieri, 2018). These correlations are consistent with expectations, as the NNAT would be expected to correlate less strongly with verbal abilities, due to its intention of measuring abilities not impacted by language (Naglieri, 2018). Therefore, it can be concluded that the NNAT and the OLSAT measure similar constructs of general and nonverbal general ability.

The Cognitive Abilities Test (CogAT)

The CogAT Form 7 (Lohman, 2012) is a group ability test that measures reasoning across three domains: Verbal, Quantitative and Nonverbal. During test administration, a test proctor guides a group of students through the test. First, the test proctor reads aloud general testing directions. At the beginning of each subtest, the test proctor completes two sample questions with the group. These sample questions demonstrate how to answer the upcoming questions. After the sample questions are complete, and students have the opportunity to ask questions, students then work independently on similar test questions of increasing complexity. For younger students (grades K-2), the test proctor reads every question to the group of students. For older students, students are given a time-limit for each subtest and are unable to go-back to prior subtests once the time-limit expires.

The CogAT used the split-half method for assessing reliability (Lohman, 2012; Warne, 2015). The use of the split-half method does not require two test administrations. Instead, the test is split into two halves and each half is compared to the other (Carmines & Zeller, 2011). Essentially, each half of the test mimics an alternative form (Carmines & Zeller, 2011). The CogAT's Research and Development Guide (Lohman, 2012) reports that the split half reliability of each independent battery across all levels of the CogAT are between .80 and .92. The split-half reliability for the composite score across all levels of the CogAT fell between .91 and .92 (Lohman, 2012; Warne, 2015). Generally, the CogAT's reliability values increased in the levels of the test meant for older children. However, this pattern is rather typical of measures of cognitive abilities (Warne, 2015) and is even explicitly expected in the Standards for Educational and Psychological Measurement (American Educational Research Association [AERA] et al., 1999).

Validation studies of the CogAT took similar measures in comparison to validation studies of the NNAT. Students' performance on the CogAT was also compared to their performance on NNAT2 (Naglieri, 2008) and the Wechsler Intelligence Scale for Children, Fourth Edition (WISC-IV). All three batteries of the CogAT were found to have correlations of at least r=.5 with scores obtained on the NNAT2 (Lohman, 2012; Warne, 2015). Furthermore, composite scores on the CogAT were found to positively correlate (r=.76) with Full-Scale IQ scores on the WISC-IV (Lohman, 2012; Warne, 2015). The results of these correlational studies support that CogAT is a valid measure of general ability (Lohman, 2012).

The Verbal Battery. The Verbal Battery of the CogAT consists of three tasks: Picture/Verbal Analogies, Sentence Completion, and Picture/Verbal Classification (Lohman, 2012).

In the Picture Analogies subtest (used for students in kindergarten through second grade), each question provides a 2x2 matrix with three pictures and one empty cell. Students must determine relationships between pictures in the matrix and apply this relationship to choose a picture that belongs in the missing cell. According to the CogAT's Research and Development Guide, most of the items in the Picture Analogies subtest were based on one of four different types of relationships. First, some items relationships were based in similarity of appearance (A looks like B). Second, items are based off action relationships, where A is doing the same thing as B. Third, some items require students to identify a transformation (A becomes B). Lastly, other items require students to understand the function of the pictures (A is used for B) when identifying the correct answer (Lohman, 2012).The Verbal Analogies subtest (used in forms for third through twelfth graders) replaces pictures with words, and words are presented in an A→B: C→ ? format instead of a matrix (Lohman, 2012). Students must determine how a pair of words go together and use this information to choose a word that completes a second pair of words in the same way. Eight types of relationships are present in these items: class inclusion (A is a member of B), part/whole (A is a part of B), Contrast/Opposite (A is the opposite of B), similar/synonym (A is similar to B), Attribute (B is an attribute of A), causation (A causes B), temporally related (B occurs during A) and case relationships (A creates B) (Lohman, 2012).

In the Sentence Completion subtest, students either listen to or read a sentience or question, and then selects best picture/word that completes the sentence or answers the question. Items are read out loud to students taking the picture-based forms of the CogAT's. On these forms, the sentence completion items were designed to test the following aspects of verbal reasoning: general verbal knowledge, deductive and inferential reasoning, vocabulary, and contra-factual thinking (Lohman, 2012). In non-picture-based forms, more complex items require a "nuanced understanding of commonly used words" (Lohman, 2012, p. 20), such as the difference between big and tall, or understanding of relatively common but abstract ideas (Lohman, 2012).

In the Picture Classification subtest, children examine three pictures and determine how these pictures are alike. Then, students select a picture that belongs in the same group as the pictures in the top row. Pictures are replaced with words in higher levels of the CogAT (Verbal Classification) (Lohman, 2012). In the picture forms, pictures are classified on the premise of the following relationships: visual matching, association, function, and shared feature (Lohman, 2012).

The Quantitative Battery. The Quantitative Battery of the CogAT consists of three tasks: Number Analogies, Number Puzzles, and Number Series (Lohman, 2012).

On lower levels of the CogAT, the Number Analogies subtest is composed of matrix items similar to those of Picture Analogies. However, the relationships depicted in the matrices require quantitative conceptual reasoning. In higher levels of the CogAT, number analogies are presented in pairs of numbers. Students identify a transformation in the first pair of numbers, and apply the same transformation to complete an incomplete pair.

In the Number Puzzles subtest students are presented with equations that have at least one missing part. If one number is missing, student must select a number that completes the puzzle. If two numbers are missing students must solve for each missing part, then solve the solution.

In lower forms of the CogAT, the Number Series subtest includes items that show several strings of beads. Students must recognize a pattern and choose the string of beads that comes next in the pattern. In higher level forms of the CogAT, students are simply given a series of numbers that have a pattern and must chose the number that comes next (Lohman, 2012).

The Nonverbal Battery. The Nonverbal Battery of the CogAT consists of three tasks: Figure Matrices, Figure Classification, and Paper Folding (Lohman, 2012). In the Figure Matrices subtest, each question shows a matrix that requires students to infer then apply a simple rule. Students must determine the relationship and select the figure that completes the matrix based upon the determined relationship. In the Picture Classification subtest, children must determine how three figures are similar. Then, they select the figure that is most like the first three figures. Major features of the figures in each item include the figures shape, size, shading pattern, symmetry and the number of distinct elements in each figure. The Paper Folding subtest requires that students imagine what happens to a piece of paper that is folded, cut in some way, and then unfolded. Students must select the answer choice that depicts how the paper will look when it is unfolded (Lohman, 2012).

Academic Achievement

The Weschler Individual Achievement Test, Third Edition (WIAT-III) (Weschler, 2009) is an individually administered achievement test that can be administered to children between preschool to twelfth grade. Administration of the WIAT-III provides information regarding a students' reading, mathematics, oral-language and writing abilities in comparison to their same-aged peers (age norms) or children within the same trimester of a particular grade level (grade norms). This dissertation utilized age-based norms, as age-based norms were utilized by participants' schools and the university training clinic when administering the CogAT and the NNAT.

The WIAT-III (Weschler, 2009) is considered a reliable measure of academic achievement. All subtest reliability coefficients fell between .83 and .97 and were calculated using either a split-half reliability method (subtests with individual items) or a test-retest method (subtests that do not have item-level responses). Furthermore, interrater reliability was also investigated in the development of the WIAT-III. For aspects of the test that are more objectively scored, interrater agreement was very high (98%-99%). On some of the more subjective subtests (i.e. Reading Comprehension, Oral Expression) interrater reliability was found to range between 91% and 99% (Weschler, 2009).

The WIAT-III (Weschler, 2009) is considered a valid measure of academic achievement. Intercorrelation studies between subtests of the WIAT-III were conducted by its test developers. Subtests that are considered in the same academic domain had moderate to higher correlations with each other (*r* =.41-.93). Furthermore, subtests in different, but related academics, also were found to have moderate to high correlations (Weschler, 2009). The WIAT-III was also compared to its previous edition, the Weschler Individual Achievement Test – Second Edition (WIAT-II; Weschler, 2001). Correlations between the WIAT-III and the WIAT-II were consistent across versions. Some subtests that had been heavily revised had lower correlations (Weschler, 2009).

During data collection, eight subtests were administered to obtain information regarding each participants achievement in reading, mathematics, and oral language. The administration of these eight subtest yields four index scores: Total Reading, Basic Reading, Mathematics and Oral-Language. However, because this study was only concerned with Reading and Mathematics achievement, participants performance on the Oral Language index, and the subtests that comprise it, were not utilized.

The subtests of the WIAT-III administered to participants yield six scores that represent reading achievement. Two of these scores are index scores (Total Reading and Basic Reading). Four scores are subtest-level scores (Word Decoding, Pseudoword Decoding, Reading Comprehension and Oral Reading Fluency). The Total Reading index score will be used as one outcome variable representing students' "Broad Reading" achievement. Established research on cognitive–achievement relationships typically breaks down Broad Reading achievement into two groups: Basic Reading Skills and Reading Comprehension (McGrew & Wendling, 2010). For this reason, this study also uses the WIAT-III's Basic Reading index score and Reading Comprehension subtest score in data analysis. Thus, this study has three outcome variables pertaining to students' reading achievement (Broad Reading, Basic Reading, and Reading Comprehension).

The Total Reading index score ("Broad Reading") is based upon a student's performance on four subtests (Reading Comprehension, Oral Reading Fluency, Word Reading, and Pseudoword Decoding). Whereas the Basic Reading index score is only based upon a student's performance on the Word Reading and Pseudoword Decoding subtests. The Word Reading subtest measures a child's speed and accuracy of decontextualized word recognition. The Pseudoword Decoding subtest measures a child's ability to decode nonsense words. The Reading Comprehension subtest measures untimed reading comprehension of various types of texts including fictional stories, informational text, advertisement and how-to passages. The Oral Reading Fluency subtest measures a child's accuracy and fluency of contextualized oral reading.

The subtests of the WIAT-III administered to participants yield three scores that represent mathematics achievement: one index score (Mathematics) and two subtest-level scores (Math Problem Solving and Numerical Operations) The proposed study is interested in "Broad Mathematic" achievement, as well as narrow the narrow skill sets measured at the sub-test level. Therefore, all three mathematics scores will be used as outcome variables that represent mathematics achievement; with the mathematics index score being referred to as "Broad Mathematics achievement," and the Numerical Operations subtest being referred to as "Math Calculation skills". This is consistent with prior research investigating cognitive–achievement relations, which has investigated both broad and narrow skills in mathematics achievement (McGrew & Wendling, 2010).

The Mathematics index score in based upon a child's performance on two subtests (Math Problem Solving, and Numerical Operations). The Math Problem Solving subtest measures untimed math-problem solving skills in the following domains: basic concepts, everyday applications (word problems). The Numerical Operations subtest measures untimed written math calculation skills.

Background History Questionnaire.

Parents were asked to complete a background history questionnaire. This questionnaire provided participants' demographic information. Namely, this questionnaire asked parents to report their child's racial identity, their family income, and if their child qualified for free or reduced price lunch in their public schools. Research supports vast socio-economic and racial group differences in academic achievement in the United States (Condron, 2009). For this reason, all analyses will control for participants' parent-reported student racial identity and family income. Family income was coded based on whether a participants family qualified for free or reduced-price lunch within their local school district.

Additionally, this questionnaire asked parents' questions regarding their students educational history, including exposure to gifted programming. Participants varied on the type of instruction they had received. Some participants were enrolled in a local school districts' tiered gifted education program. This program provides students with an accelerated curriculum. Because achievement is highly related to instruction, control variables will be included in relevant equations to account for differences in instruction. Dummy variables were utilized to control for the effects of full-time and part-time gifted educational services (in reference to no gifted educational services). An additional dummy variable was also utilized for participants who were homeschooled or attended private school.

Review of Academic Records

Parents were also asked to provide educational records. Namely, official score reports of their children's' group-ability testing with the NNAT and the CogAT. These official score reports detailed when group-ability testing took place for each individual participant. The amount of time that occurs between group-ability testing (predictor) and individual achievement testing (outcome) varied between participants. For example, some participants were administered the group-ability tests several years before follow-up achievement testing whereas others were administered both assessments within the same year. For this reason, a control variable was added to the analyses that represented the amount of time between group-ability testing and individual achievement testing. Additionally, although most participants took the NNAT and CogAT in first and second grade respectively, some took these group-ability tests at other times. In order to account for these differences, another control variable will be utilized to represent the amount of time after first or second grade each test was taken.

CHAPTER 7: DATA ANALYSIS

The data analysis plan for each research question is listed below. In efforts to reduce redundancy, the following equations are written with “Achievement” as the outcome variables. In all analyses, there are six “achievement” outcomes: Broad Reading, Basic Reading, Reading Comprehension, Broad Mathematics, Math Problem Solving, and Math Calculation.

1. Are there group-differences in achievement for students who have received “gifted” programming in the form of above-grade level instruction and students who follow a general curriculum?

   Group differences between students in gifted programming will be explored using one-way ANOVA. Comparisons will be made between three gifted programming groups: a full-time gifted programming group, a part-time gifted programming group, and a group that receives no gifted programming. Participants who were homeschooled or enrolled in a private school will be excluded from the analysis. Post-hoc testing will be utilized to explore any significant differences.

2. How do students' CogAT Composite score and NNAT score relate to reading and mathematics achievement? Are there differences in test–achievement relationships across these different academic domains?

A series of regression analyses will be utilized to explore this research question. Each regression will use either the NNAT composite score or the CogAT composite score as a predictor of academic achievement while accounting for control variables. In equations that use the CogAT Composite Score, time since 2nd grade will be used as a control variable (since most participants took the CogAT in the second grade). In equations that use the NNAT score, time since 1st grade will be used a control variable (since most participants took the NNAT in the first grade). Because the proposed study will explore 3 reading-related achievement outcomes (Broad Reading, Basic Reading, Reading Comprehension) and 3 mathematic-related achievement outcomes (Broad Mathematics, Math Problem Solving, Math Calculation), a total of 12 regression equations will be used to investigate this research question. Equation 1 presents a generalized version of the regressions used to answer this question.

(1) $\text{Achievement} = \beta_0 + \beta_1(\text{race}) + \beta_2(\text{income}) + \beta_3(\text{time since NNAT/CogAT}) + \beta_4(\text{time since 1st/2nd grade}) + \beta_5(\text{Level Gifted Programming}) + \beta_6(\text{NNAT/COGAT})$

Stepwise regression (Equations 2 and 3) will also be utilized to further explore the predictive abilities of the NNAT and CogAT Composite. Because most participants in the sample have taken the NNAT prior to taking the CogAT, the NNAT score will be added to the model first. Then the CogAT Composite score will be added in a second step. The results of this second step will provide more direct information regarding the predictive abilities of the NNAT in comparison to the CogAT Composite score.

(2) Step 1: Academic Achievement = $\beta_0 + \beta_1$(race) + β_2(income) + β_3(time since) + β_4(time since 1st grade) + β_5 (Level Gifted Programming) + β_6 (NNAT)

(3) Step 2: Academic Achievement = $\beta_0 + \beta_1$(race) + β_2(income) + β_3(time since NNAT) + β_4(time since 1st grade) + β_5 (Level Gifted Programming) + β_6(NNAT) + β_7 (time since CogAT) + β_8 (time since 2nd grade) + β_8 (CogAT Composite)

3. How well does the CogAT Composite Score relate to academic achievement in comparison to CogAT battery scores (Verbal, Quantitative and Nonverbal)? Which combinations of scores from the CogAT are most relevant to students' academic achievement?

This research question can be seen as an extension of the second research question; and will be explored in a similar manner. Each battery of the CogAT (Verbal, Quantitative, and Nonverbal) will be regressed onto each domain of academic achievement (Total Reading, Basic Reading, Mathematics, Math Problem Solving, and Numerical Operations). In total, 18 regression equations will be utilized to answer this question. Equation 4 presents a generalized version of the regressions used for each battery of the CogAT.

(4) Achievement =$\beta_0 + \beta_1$(race) + β_2(income) + β_3(time since CogAT) + β_4(time since 2nd grade) + β_5 (Level Gifted Programming) + β_6 (CogAT Verbal/Quantitative/Nonverbal Battery)

In order to explore the second part of this research question, each battery score will be tested alone and in combination with other batteries found to be predictive of each outcome. Stepwise regression analysis will be used to determine if models with multiple battery scores are better predictors of academic achievement, in comparison to models with one battery score.

Depending on the results of the previous regression analyses, a hierarchy of predictor variables (Battery Scores) will be determined for each academic achievement area. This hierarchy will be established based upon the extent to which each battery is found to significantly predict each of the academic outcomes on its own. For each outcome, the strongest predictor will be added in the first step of the regression. Then, other battery scores will be added based upon their observed predictive abilities. If the addition of the score explains a significant amount of variance in achievement, above and beyond those already added into the model then it will be added. However, if addition of the score does not explain a significant amount of variance above and beyond the already-added scores it will be removed from the model. This regression will be compared to Equation 1 and 4's models to determine whether which models best predict academic achievement.

4. Does the Nonverbal Battery of the CogAT provide significant predictive value above and beyond students' NNAT scores when predicting Reading and Mathematics Achievement?

This research question will be investigated using stepwise regression. Because most participants in the sample have taken the NNAT prior to taking the CogAT, a participants NNAT score will be added to the model first. Then, a child's Nonverbal score from the CogAT will be added in another step. If the second model explains a statistically significantly higher amount of variance in student achievement, then we can conclude that the CogAT's nonverbal battery does provide predictive value. The regression models used will be:

(5) Step 1: Academic Achievement = $\beta_0 + \beta_1$(race) + β_2(income) + β_3(time since) + β_4(time since 1st grade) + β_5 (Level Gifted Programming) + β_6 (NNAT)

(6) Step 2: Academic Achievement = $\beta_0 + \beta_1$(race) + β_2(income) + β_3(time since NNAT) + β_4(time since 1st grade) + β_5 (Level Gifted Programming) + β_6(NNAT) + β_7 (time since CogAT) + β_8 (time since 2nd grade) + β_8 (Nonverbal CogAT Battery)

5. Are the relationships between the batteries of the CogAT, the NNAT and individual academic achievement moderated by the age of the student when achievement testing took place?

To answer this question, moderation regression analysis will be used. A interaction term will be added to each of each of the 5scores (NNAT, CogAT Composite, CogAT Verbal, CogAT Quantitative and CogAT Nonverbal) regressions for each of the 6 outcomes. Therefore, a total of 30 interactions will be tested. Equation 7 presents the general regression equation testing this interaction.

(7) Achievement = $\beta_0 + \beta_1$(race) + β_2(income) + β_3(time since test) + β_4(time since grade) + β_5 (Level Gifted Programming) + β_6 (NNAT/CogAT Composite/CogAT Battery) + β_7 (age) + β_8 (NNAT/CogAT*Age)

If the regression coefficient β_8 is found to be significant, then age acts as a moderator on the relationship between the test score (NNAT, CogAT Composite, CogAT battery) and academic achievement.

CHAPTER 8: RESULTS

All 75 participants provided NNAT scores. 74 participants provided CogAT scores. As shown in Table 1, the mean NNAT score of participants in the sample was 121.6 (*SD* 15.5). The mean CogAT Composite score of participants in the sample was 124.3 (*SD* 10.1) with Verbal, Quantitative and Nonverbal battery scores of 118.9 (*SD* 8.9), 123 (*SD* 11.1), and 122.4 (SD11.1) respectively.

Table 1 Descriptive Statistics of NNAT and CogAT Scores

Test Score	n	Min.	Max.	Mean	*SD*
NNAT	75	92	160	121.6	15.5
CogAT Composite	74	97	140	124.3	10.1
CogAT Verbal	74	102	139	118.9	8.9
CogAT Quantitative	74	92	145	123.0	11.1
CogAT Nonverbal	74	92	143	122.4	11.1

All 75 participants were administered select subtests of the WIAT-III. As shown in Table 2, the mean Total Reading index score of participants in the sample was 117.5 (*SD* 12.9). The mean Basic Reading index score of participants in the sample was 117.0 (*SD* 14.9). The mean score on the Reading Comprehension subtest of participants in the sample was 113.3 (*SD* 13.1). The mean score on the Mathematics index of participants was 122.8 (*SD* 16.9), with the Math Problem Solving and Numerical Operations subtests having mean scores of 121.9 (*SD* 15.2) and 121.0 (*SD* 20.9) respectively.

Table 2 Descriptive Statistics of Academic Achievement

WIAT-III Score	n	Min.	Max.	Mean	*SD*
Total Reading	75	86	142	117.5	12.9
Basic Reading	75	87	152	117.0	14.9
Reading Comprehension	75	89	160	113.3	13.1
Mathematics	75	87	160	122.8	16.9
Math Problem Solving	75	85	160	121.9	15.2
Numerical Operations	75	86	160	121.0	20.9

Students' CogAT Composite Scores and NNAT Scores were found to be significantly correlated (*r*=.56). Students NNAT scores were significantly correlated with all three batteries of the CogAT. Out of the battery scores, the NNAT scores were least correlated with the CogAT Verbal Battery Score(*r*=.31), and most correlated with the CogAT's Nonverbal Battery (*r*=.49). Correlations across the NNAT and CogAT are presented in Table 3.

Table 3 Correlations between NNAT and CogAT Scores

Test Score	1	2	3	4	5
1. NNAT		.56**	.31**	.48**	.49**
2. CogAT Composite			.66**	.84**	.81**
3. CogAT Verbal				.34**	.39**
4. CogAT Quantitative					.51**
5. CogAT Nonverbal					

* *p* <.05, ** *p* <.01

Research Question #1

The first research question investigated whether group-differences in academic achievement were present due the level of gifted programming a student receives. Of the

75 participants, 27 (36.0%) did not receive any gifted programming, 19 (24.0%) received part-time gifted programming, 21 (29.3%) received full-time gifted programming. The remaining 8 (10.6%) participants were either homeschooled or enrolled in a private school and were, thus, dropped from these analyses. Descriptive statistics of each outcome by group are presented in Table 4.

Table 4 Academic Achievement by Gifted Programming Group

	None		Part-Time		Full-Time	
Outcome	*m*	*SD*	*m*	*SD*	*m*	*SD*
Broad Reading	111.96	12.80	120.05	10.85	120.52	10.41
Basic Reading	111.52	12.80	119.84	13.06	117.48	12.84
Reading Comprehension	109.59	10.41	112.26	11.58	120.20	15.80
Broad Mathematics	113.30	10.67	124.37	15.04	130.38	15.02
Math Problem Solving	115.59	9.98	124.11	12.86	126.33	16.53
Math Calculation	109.40	12.19	121.37	20.00	130.66	19.12

A one-way ANOVA determined significant differences in Broad Reading ($F(2,64) = 4.20$, $p < .05$), Reading Comprehension ($F(2,64) = 4.32$, $p < .05$), Broad Mathematics ($F(2,64) = 10.05$, $p < .001$), Math Problem Solving ($F(2,64) = 4.53$, $p < .05$) and Math Calculation ($F(2,64) = 9.46$, $p < .05$) achievement between groups. There were no significant differences in Basic Reading Achievement across groups ($F(2,62) = 2.60$, $p > .05$).

Broad Reading

Levene's test showed that the variances of Broad Reading Achievement between groups were not different ($F(2,64) = .91$, $p = .41$). A post-hoc Tukey test revealed that students who received full-time gifted programming had significantly higher Broad

Reading achievement in comparison to students who received no gifted programming On average, students who had received full-time gifted programming scored 8.56 points higher on the WIAT-III's Total Reading index. There was no significant difference in Broad Reading achievement between student's who received no gifted programming and those who received part-time gifted programming. Nor was there a significant difference in Broad Reading between students who received part-time gifted programming and students who received full-time gifted programming.

Reading Comprehension

Levene's test showed that the variances of Reading Comprehension abilities between groups were not different ($F(2,64) = 1.64$, $p = .20$). A post-hoc Tukey test revealed that students who received full-time gifted programming performed significantly better on the WIAT-III's Reading Comprehension subtest in comparison to students who received no gifted programming. On average, students who had received full-time gifted programming scored 10.60 points higher on the WIAT-III's Reading Comprehension index than students who received no gifted programming. There was no significant difference in Reading Comprehension between student's who received no gifted programming and those who received part-time gifted programming. Nor was there a significant difference in Reading Comprehension between students who received part-time gifted programming and students who received full-time gifted programming.

Broad Mathematics

Levene's test showed that the variances of Broad Mathematic achievement between groups were not different ($F(2,64) = 1.33$, $p = .27$). A post-hoc Tukey test revealed that students who received no gifted programming had significantly lower Broad Mathematics achievement in comparison to students who had received part-time and full-time gifted programming. On average, students who had received no gifted programming performed 11.01 points lower than those enrolled in part-time gifted programming and 17.08 points lower than those enrolled in full-time gifted programming. However, there were no significant differences in Broad Mathematic achievement between the full-time and part-time groups.

Math Problem Solving

Levene's test showed that the variances of Math Problem Solving abilities between groups were not different ($F(2,64) = 1.05$, $p = .36$). A post-hoc Tukey test revealed that students who received no gifted programming had significantly lower Math Problem Solving abilities compared to students who received full-time gifted programming. There was no significant difference in Math Problem Solving between student's who received no gifted programming and those who received part-time gifted programming. Nor was there a significant difference in Math Problem Solving abilities between students who received part-time gifted programming and students who received full-time gifted programming.

Math Calculation

The assumption of equal variances was not met between groups on the Math Calculation outcome variable. Therefore, the Games-Howell post-hoc test was utilized to further investigate the differences between groups. This test determined a significant difference between students who had received full-time gifted programming and those who received no gifted programming. On average, students who had received full-time gifted programming scored 21.26 points higher on the WIAT-III's Numerical Operations subtest. There was no significant difference in Math Calculation between student's who received no gifted programming and those who received part-time gifted programming. Nor was there a significant difference in Math Calculation between students who received part-time gifted programming and students who received full-time gifted programming.

Research Question #2

The second research question investigated the predictive abilities of the NNAT score and the CogAT Composite score on reading and mathematics achievement. Regressions for reading achievement are summarized in Table 5. Regressions for mathematics achievement are summarized in Table 6.

Table 5 Reading Achievement Regressions

	Broad Reading			Basic Reading			Reading Comp.		
Test Score	*B*	*β*	*t*	*B*	*β*	*t*	*B*	*β*	*t*
NNAT	.30	.37	2.82**	.43	.46	3.50**	.17	.20	1.47
CogAT Composite	.49	.40	3.42**	.64	.44	3.74**	.23	.18	1.43
CogAT Verbal	.31	.22	1.81	.35	.22	1.71	.18	.12	1.00
CogAT Quantitative	.42	.37	3.31**	.53	.40	3.54**	.26	.23	1.84
CogAT Nonverbal	.37	.33	2.56*	.48	.36	2.83**	.12	.10	.76

* $p < .05$, ** $p < .01$

Table 6 Mathematics Achievement Regressions

	Broad Mathematics			Math Prob. Solving			Math Calculation		
Test Score	*B*	*β*	*t*	*B*	*β*	*t*	*B*	*β*	*t*
NNAT	.55	.51	4.38**	.45	.47	3.85**	.59	.45	3.71**
CogAT Composite	.80	.48	4.51**	.71	.47	4.37**	.76	.37	3.32**
CogAT Verbal	.45	.23	2.05*	.59	.30	2.65**	.29	.13	1.07
CogAT Quantitative	.63	.41	3.92**	.59	.40	3.74**	.61	.32	2.98*
CogAT Nonverbal	.61	.40	3.45**	.50	.36	3.03**	.63	.34	2.82**

* p <.05, ** p <.01

NNAT Score

Reading

A multiple regression was carried out to investigate the extent to which the NNAT score significantly predicted students' Broad Reading achievement after accounting for control variables. The results of the regression indicated that the model explained 29.7% of the variance in broad reading ($F(10,63) = 2.67$, $p = .009$). Students' NNAT scores were found to significantly predict Broad Reading achievement (B=.30, β=.37, p <.01). This indicates that on average, a score increase of one point on the NNAT translates to a .30 increase on the WIAT-III's Total Reading index.

The results of the regression predicting Basic Reading indicated that the model explained 29.5% of the variance ($F(10,63) = 2.64$, $p = .009$). Students' NNAT scores were found to significantly predict Basic Reading (B=.43, β=.46, p <.01).

The results of the regression predicting Reading Comprehension indicated that the model explained 23.0% of the variance. The NNAT score did not significantly predict Reading Comprehension abilities (B=.17, β=.20, p >.05). Further, this model was not found to be a significant predictor of Reading Comprehension abilities, ($F(10, 63) = 1.881$, $p = .07$).

Mathematics

A multiple regression was carried out to investigate the extent to which the NNAT score significantly predicted students' Broad Mathematics achievement after accounting for control variables. The results of the regression indicated that the model explained 43.5% of the variance, and that the model was a significant predictor of Broad Mathematics achievement, ($F(10,63) = 4.85$, $p < .001$). The NNAT score was found to significantly predict Broad Mathematics achievement (B=.55, β=.51, p <.01). This indicates that on average, a score increase of one point on the NNAT translates to about ½ a point increase on the WIAT-III's Mathematics index.

A The results of the regression predicting Math Problem Solving abilities indicated that the model explained 39.2% of the variance, and that the model was a significant predictor of Math Problem Solving abilities, ($F(10,63) = 4.10$, p <.001). The NNAT score was found to significantly predict Math Problem Solving abilities (B=.45, β=.47, $p < .01$).

The results of the regression predicting Math Calculation abilities indicated that the model explained 39.8% of the variance, and that the model was a significant predictor, ($F(10,63) = 4.17$, p <.001). Students' NNAT scores were found to significantly predict math calculation abilities (B=.59, β=.45, p <.01).

CogAT Composite Score

Reading

A multiple regression was carried out to investigate the extent to which the CogAT Composite score could significantly predict students' Broad Reading achievement after accounting for control variables. The results of the regression indicated that the model explained 32.3% of the variance ($F(10,62) = 2.96$, $p = .004$). The CogAT Composite score was found to significantly predict students' Broad Reading achievement ($B=.49$, $\beta=.40$, $p <.01$). This indicates that on average, a score increase of one point on the CogAT Composite translates to a .49 increase on the WIAT-III's Total Reading index.

The results of the regression predicting Basic Reading achievement indicated that the model explained 30.9% of the variance, and that the model was a significant predictor of Basic Reading achievement ($F(10,62) = 2.77$, p = .007). The CogAT Composite score was found to significantly predict Basic Reading Achievement ($B=.64$, $\beta=.44$, $p <.01$).

The results of the regression predicting Reading Comprehension abilities indicated that the model explained 23.6% of the variance. The CogAT Composite score did not significantly predict Reading Comprehension abilities ($B=.23$, $\beta=.18$, $p >.05$). Further, this model was not found to be a significant predictor of Reading Comprehension abilities ($F(10, 62) = 1.91$, $p = .060$).

Mathematics

A multiple regression was carried out to investigate the extent to which the CogAT Composite score could significantly predict students' Broad Mathematics achievement after accounting for control variables. The results of the regression indicated that the model explained 44.8% of the variance (F(10,62) = 5.02, p <.001). The CogAT Composite score was found to significantly predict Broad Mathematics achievement (B=.80, β=.48, p <.01). This indicates that, on average, a score increase of one point on the CogAT Composite translates to an .80 increase on the WIAT-III's Mathematics index.

The results of the regression predicting Math Problem Solving abilities indicated that the model explained 43.0% of the variance (F(10,62) = 4.68, p = <.001). The CogAT Composite score was found to significantly predict Math Problem Solving abilities (B=.71, β=.47, p <.01).

The results of the regression predicting Math Calculation abilities indicated that the model explained 37.9% of the variance, (F(10,62) = 3.78, p = .001). Students' CogAT Composite scores were found to significantly predict Math Calculation abilities (B=.76, β=.37, p <.01).

NNAT vs. CogAT Composite

Stepwise regression was utilized to determine whether the CogAT Composite could provide significant predictive value above and beyond students' NNAT scores. The CogAT Composite score was only added to the models in which the NNAT score was found to be a significant predictor (Broad Reading, Basic Reading, Broad Mathematics, Math Problem Solving, and Math Calculation). The results are summarized in Table 7 (reading achievement domains) and Table 8 (mathematic achievement domains).

Table 7 Stepwise Regressions NNAT and CogAT Composite (Reading Achievement)

		Broad Reading			Basic Reading		
Step	Predictor	*B*	β	*t*	*B*	β	*t*
1	NNAT	.29	.35	2.68**	.41	.44	3.33**
2	NNAT	.17	.22	1.46	.30	.32	2.17*
	CogAT Composite	.39	.32	2.42*	.46	.32	2.45*

* p <.05, ** p <.01

In stage one of the multiple regression predicting Broad Reading achievement, the NNAT score significantly contributed (B =.29, β = .35, p <.01) to a model that explained 28.4% of the variance, $F(10, 62) = 2.454$, p=.02. Introducing the CogAT's Composite score in stage 2 explained an additional 7.0% of the variance. This R^2 change was found to be significant (p=.05). In this second model, the CogAT Composite score (B =.39, β = .32, p <.05) significantly predicted Broad Reading achievement. However, the NNAT score (B =.17, β = .22, p >.05) no longer predicted Broad Reading Achievement.

In stage one of the multiple regression predicting Basic Reading achievement, the NNAT score significantly contributed (B =.41, β = .44, p <.01) to a model that explained 28.5% of the variance, $F(10, 62) = 2.47$, p=.02. Introducing the CogAT's Composite score in stage 2 explained an additional 8.2% of the variance. This R^2 change was found to be significant (p=.03). In this second model, both the CogAT Composite Score and the NNAT Score significantly predicted Basic Reading achievement ($p<.05$).

Table 8 Stepwise Regressions NNAT and CogAT Composite (Mathematics Achievement)

		Broad Mathematics			Math Prob. Solving			Math Calculation		
Step	Predictor	*B*	*β*	*t*	*B*	*β*	*t*	*B*	*β*	*t*
1	NNAT	.57	.53	4.58**	.48	.48	4.00**	.62	.46	3.84**
2	NNAT	.43	.39	3.08**	.31	.32	2.40*	.56	.42	3.07**
	CogAT Composite	.54	.23	2.86**	.52	.34	2.94**	.33	.18	1.42

* p <.05, ** p <.01

In stage one of the multiple regression predicting Broad Mathematics achievement, the NNAT score significantly contributed (B =.57, β= .53, p <.001) to a model that explained 45.2% of the variance, F(10, 62) = , p= < .001. Introducing the CogAT's Composite score in stage 2 explained an additional 7.3% of the variance. This R_2 change was found to be significant (p=.01). In this second model, both the CogAT Composite Score and the NNAT Score significantly predicted Broad Mathematics achievement (p<*.01*).

In stage one of the multiple regression predicting Math Problem Solving abilities, the NNAT score significantly contributed (B =.48, β= .48, p <.01) to a model that explained 40.7% of the variance, F(10, 62) = , p= < .001. Introducing the CogAT's Composite score in stage 2 explained an additional 7.6% of the variance. This R_2 change was found to be significant (p=.02). In this second model, both the CogAT Composite Score and the NNAT Score significantly predicted Broad Mathematics achievement (p<*.01*).

In stage one of the multiple regression predicting Math Calculation skills, the NNAT score significantly contributed (*B* =.62, β = .46, *p* <.001) to a model that explained 40.9% of the variance, *F*(10, 62) = , p= < .001. Introducing the CogAT's Composite score in stage 2 explained an additional 4.3% of the variance. However, this R_2 change was not found to be significant (*p*=.11).

Research Question #3

The third research question investigates the predictive abilities of the three distinct batteries of the CogAT (Verbal, Quantitative and Nonverbal) on reading (Table 5) and mathematics achievement (Table 6).

CogAT Verbal Battery

Reading

A multiple regression was carried out to investigate the extent to which the CogAT Verbal Battery score could significantly predict students' Broad Reading achievement after accounting for control variables. The results of the regression indicated that the model explained 23.6% of the variance and was non-significant overall (*F*(10,62) = 1.91, *p* = .06). A students' CogAT Verbal Battery score was not found to be a significant predictor of the Broad Reading (*B*=.31, β=.22, *p* >.05).

The results of the regression predicting Basic Reading achievement indicated that the model explained 19.1% of the variance and was non-significant overall (*F*(10, 62) = 1.46, *p* = .18). The CogAT Verbal Battery score did not significantly predict Basic Reading achievement (*B*=.35, β=.20, *p* >.05).

The results of the regression predicting Reading Comprehension abilities indicated that the model explained 22.3% of the variance and was non-significant overall ($F(10, 62) = 1.78$, $p = .08$). The CogAT Verbal Battery score did not significantly predict Reading Comprehension abilities (B=.18, β=.12, p >.05).

Mathematics

A multiple regression was carried out to investigate the extent to which the CogAT Verbal Battery score could significantly predict students' Broad Mathematics achievement after accounting for control variables. The results of the regression indicated that the model explained 31.3% of the variance, ($F(10,62) = 2.82$, $p = .006$). The CogAT Verbal Battery was found to significantly predict Broad Mathematics achievement (B=.45, β=.24, p <.05). This indicates that on average an increase of one point on the CogAT's Verbal Battery translates to about a ½ a point increase on the WIAT-III's Mathematics index.

The results of the regression predicting Math Problem Solving abilities indicated that the model explained 33.1% of the variance ($F(10,62) = 3.06$, $p = .003$). The CogAT Verbal Battery score was found to significantly predict Math Problem Solving abilities (B=.52, β=.30, p >.01).

The results of the regression predicting Math Calculation abilities indicated that the model explained 28.1% of the variance ($F(10,62) = 2.42$, $p = .017$). However, the CogAT Verbal Battery score was not found to significantly predict Math Calculation abilities (B=.29, β=.13, p >.05).

CogAT Quantitative Battery

Reading

A multiple regression was carried out to investigate whether the CogAT Quantitative Battery score could significantly students' Broad Reading achievement after accounting for control variables. The results of the regression indicated that the model explained 31.6% of the variance ($F(10,62) = 2.86$, $p = .005$). Students' CogAT Quantitative Battery scores were found to significantly predict Broad Reading Achievement (B=.42, β=.37, p <.01). This indicates that on average, a score increase of one point on the CogAT Quantitative Battery translates to a .42 increase on the WIAT-III's Total Reading index.

The results of the regression predicting Basic Reading achievement indicated that the model explained 29.5% of the variance ($F(10,62) = 2.60$, $p = .011$). The CogAT Quantitative Battery score was found to significantly predict Basic Reading achievement (B=.53, β=.40, p <.01).

The results of the regression predicting Reading Comprehension abilities indicated that the model explained 25.2% of the variance ($F(10,62) = 2.08$, $p = .039$). However, the CogAT Quantitative Battery score did not significantly predict Reading Comprehension abilities (B=.26, β=.14, p >.05).

Mathematics

A multiple regression was carried out to investigate the extent to which the CogAT's Quantitative Battery Score could significantly predict students' Broad Mathematics achievement after accounting for control variables. The results of the regression indicated that the model explained 41.2% ($F(10,62) = 4.35$, $p = <.001$). The CogAT Quantitative Battery significantly predicted Broad Mathematics achievement ($B=.63$, $\beta=.41$, $p <.01$). This indicates that on average, a score increase of one point on the CogAT Quantitative Battery translates to an .63 increase on the WIAT-III's Mathematics index.

The results of the regression predicting Math Problem Solving abilities indicated that the model explained 39.3% ($F(10,62) = 3.99$, $p = <.001$). The CogAT Quantitative Battery score was found to significantly predict Math Problem Solving abilities ($B=.55$, $\beta=.40$, $p <.01$).

The results of the regression predicting Math Calculation abilities indicated that the model explained 35.9% ($F(10,62) = 3.48$, $p = .001$). The CogAT Quantitative Battery score was found to significantly predict Math Calculation abilities ($B=.61$, $\beta=.32$, $p <.01$).

CogAT Nonverbal Battery

Reading

A multiple regression was carried out to investigate whether the extent to which the CogAT's Nonverbal Battery Score could significantly predict a child's Broad Reading achievement after accounting for control variables. The results of the regression indicated that the model explained 27.3% of the variance ($F(10,62) = 2.33$, $p = .021$). The CogAT Nonverbal Battery score was found to significantly predict Broad Reading Achievement ($B=.37$, $\beta=.33$, $p <.05$). This indicates that on average, a score increase of one point on the CogAT Nonverbal Battery translates to an increase of .37 on the WIAT-III's Total Reading index.

The results of the regression predicting Basic Reading abilities indicated that the model explained 24.9% of the variance ($F(10,62) = 2.06$, $p = .04$). The CogAT Nonverbal Battery score was found to significantly predict Basic Reading (B=.48, β=.36, p <.01).

The results of the regression predicting Reading Comprehension abilities indicated that the model explained 21.8% of the variance, and was non-significant overall ($F(10,62) = 1.73$, $p = .09$). The CogAT Nonverbal Battery was not found to be a significant predictor of Reading Comprehension abilities (B=.12, β=.10, p >.05).

Mathematics

A multiple regression was carried out to investigate the extent to which the CogAT's Nonverbal Battery Score could significantly predict a child's Broad Mathematics achievement accounting for control variables. The results of the regression indicated that the model explained 38.4% of the variance ($F(10,62) = 3.87$, p = < .001). The Nonverbal Battery score was found to significantly predict Broad Mathematics achievement ($B = .61$, β=.40, p <.01). This indicates that on average, a score increase of one point on the Nonverbal Battery translates to a .61 increase on the WIAT-III's Mathematics index.

The results of the regression predicting Math Problem Solving abilities indicated that the model explained 35.1% of the variance ($F(10,62) = 3.35$, $p = .002$). The student's Nonverbal Battery score was found to significantly predict Math Problem Solving abilities (*B*= .50, β=.36, p <.01).

The results of the regression predicting Math Calculation abilities indicated that the model explained 35.1% of the variance ($F(10,62) = 3.36$, $p = .002$). The Nonverbal Battery score was found to significantly contribute predict Math Calculation abilities (*B* = .63, β=.34, p <.01).

Battery Score Models

The third research question utilized stepwise regression to investigate the extent to which models with multiple battery scores could predict academic achievement in reading and mathematics. For each outcome, a hierarchy was established based on the each battery's standardized coefficients for each outcome variable (Tables 5 and 6).

When predicting Broad Reading, Basic Reading, and Broad Mathematics the Quantitative Battery score was entered in the first step of the regression equation. The results of these regressions are presented in Table 9. The Nonverbal Battery was introduced in the second step for each of these outcomes. The introduction of the Nonverbal Battery in prediction of Broad Reading achievement explained 3.1% more variance than the Quantitative Battery alone. However, this change in R^2 was not significant (p=.07). The introduction of the Nonverbal Battery in prediction of Basic Reading achievement explained 2.0% more variance than the Quantitative Battery alone. However, this change in R^2 was not significant (p=.19). The introduction of the Nonverbal Battery in prediction of Broad Mathematics achievement explained 1.5% more variance than the Quantitative Battery alone. However, this change in R^2 was not significant (p=.25). In all three of these models, the CogAT's Nonverbal Battery was no longer a significant predictor of the outcome.

Table 9 Stepwise Regressions Battery Score Models (Broad Reading, Basic Reading, Broad Mathematics)

		Broad Reading			Basic Reading			Broad Mathematics		
Step	Predictor	*B*	*β*		*B*	*β*	*t*	*B*	*β*	*t*
1	CogAT Quantitative	.42	.37	*t*	.53	.40	3.54**	.63	.16	3.92**
2	CogAT Quantitative	.34	.30	.2.28*	.42	.32	1.33*	.46	.30	2.54*
	CogAT Nonverbal	.19	.16	1.16	.25	.19	1.33	.36	.20	1.84

* p <.05, ** p <.01

When predicting Math Problem Solving abilities the Quantitative Battery was entered in the first step of the regression. The Verbal Battery was introduced in the second step. The introduction of the Verbal Battery explained 2.1% more variance than the Quantitative Battery alone. However, this change in R^2 was not significant (p=.15) and the Verbal Battery was no longer a significant predictor of Math Problem Solving. The Verbal Battery was dropped, and the Nonverbal Battery was introduced in step 2b. The introduction of the Nonverbal Battery explained 2.1% more variance than the Quantitative Battery alone. However, this change in R^2 was not significant (p=.15) and the Nonverbal Battery was no longer a significant predictor of Math Problem Solving. The results of these regressions are presented in Table 10.

Table 10 Stepwise Regressions Battery Score Models (Math Problem Solving)

		Math Problem Solving		
Step	Predictor	*B*	*β*	*t*
1	CogAT Quantitative	.55	.40	2.74**
2a	CogAT Quantitative	.46	.33	2.92**
	CogAT Verbal	.30	.17	1.476
2b	CogAT Quantitative	.43	.31	2.52**
	CogAT Nonverbal	.27	.19	1.47

* p <.05, ** p <.01

When predicting Math Calculation skills the Nonverbal Battery was entered in the first step of the regression. The Quantitative Battery was introduced in the second step. The introduction of the Quantitative Battery explained 3.3% more variance than the Nonverbal Battery alone. However, this change in R_2 was not significant (p=.07), and both the Nonverbal and Quantitative batteries no longer predicted Math Calculation skills. The results of these regressions are presented in Table 11.

Table 11 Stepwise Regressions Battery Score Models (Math Calculation)

		Math Calculation		
Step	Predictor	*B*	*β*	*t*
1	CogAT Nonverbal	.63	.34	2.82**
2	CogAT Nonverbal	.40	.21	1.57
	CogAT Quantitative	.43	.23	1.82

* p <.05, ** p <.01

Research Question #4

The third research question aimed to explore whether the CogAT's Nonverbal Battery explained variance in students' achievement scores above and beyond the variance explained by the NNAT score. The Nonverbal Battery was only added to the models in which the NNAT score was found to be a significant predictor (Broad Reading, Basic Reading, Broad Mathematics, Math Problem Solving, and Math Calculation). The results are summarized in Table 12 (reading achievement domains) and Table 13 (mathematic achievement domains).

Table 12 Stepwise Regressions Nonverbal Measures (Reading Achievement)

		Broad Reading			Basic Reading		
Step	Predictor	*B*	β	*t*	*B*	β	*t*
1	NNAT	.29	.35	2.68**	.41	.44	3.33**
2	NNAT	.23	.28	1.89	.36	.39	2.61*
	CogAT Nonverbal	.24	.22	1.57	.28	.21	1.56

* p <.05, ** p <.01

In stage one of the multiple regression predicting Broad Reading achievement, the NNAT score significantly contributed (B =.29, β= .35, p <.01) to a model that explained 28.4% of the variance, $F(10, 62) = 2.45$, p=.015. The change in R_2 was not significant from Step 1 to Step 2 (p =.24). Furthermore, in this second model the NNAT Score (B=.23, $\beta = .22$, $p > .05$) and the CogAT's Battery score (B=.24, $\beta = .22$, $p > .05$) both no longer significantly predicted Broad Reading achievement.

In stage one of the multiple regression predicting Basic Reading achievement, the NNAT score significantly contributed (B =.41, β= .44, p = <.01) that explained 28.5% of the variance, $F(10, 62) = 2.466$, p=.02. The change in R^2 was not significant from Step 1 to Step 2 (p=.14) and the CogAT's Nonverbal Battery was no longer a significant predictor of Basic Reading Achievement (B =.28, β= .21, p = >.05).

Table 13 Stepwise Regressions Nonverbal Measures (Mathematics Achievement)

		Broad Mathematics			Math Prob. Solving			Math Calculation		
Step	Predictor	*B*	β	*t*	*B*	β	*t*	*B*	β	*t*
1	NNAT	.57	.53	4.58**	.48	.48	4.00**	.62	.46	3.84**
2	NNAT	.50	.46	3.53**	.40	.40	2.96**	.56	.42	3.07**
	CogAT Nonverbal	.34	.23	1.90	.28	.21	1.66	.33	.18	1.42

* p <.05, ** p <.01

In stage one of the multiple regression predicting Broad Mathematics achievement, the NNAT score significantly contributed (B =.57, β= .53, p = <.01) that explained 45.2% of the variance, $F(10, 62) = 5.11$, p= <.001. The change in R^2 was not significant from Step 1 to Step 2 (p=.11) and CogAT's Nonverbal Battery no longer served as a significant predictor of Broad Mathematics Achievement (B =.34, β= .23, p >.05).

In stage one of the multiple regression predicting Math Problem Solving abilities, the NNAT score significantly contributed (B =.48, β= .48, p = <.01) that explained 40.7% of the variance, $F(10, 62) = 4.26$, p= <.001. The change in R^2 was not significant from Step 1 to Step 2 (p =.24) and the CogAT's Nonverbal Battery no longer served as a significant predictor of Math Problem Solving abilities (B =.28, β = .21, p >.05).

In stage one of the multiple regression predicting Math Calculation abilities, the NNAT score significantly contributed (B =.62, β = .46, p <.01) to a model that explained 40.9% of the variance, F(10, 62) =4.28, p = <.001. Introducing the CogAT's Nonverbal Battery score in stage 2 explained an additional 3.1% of the variance. The change in R^2 was not significant from Step 1 to Step 2 and the CogAT's Nonverbal Battery no longer served as a significant predictor of Math Problem Solving abilities (B =.33, β = .18, p >.05).

These stepwise regressions utilized to answer this research question were also conducted in the opposite order. The CogAT Nonverbal was entered at Step 1 and the NNAT score was entered at step 2. In these stepwise equations, similar findings were found.

Research Question #5

The fifth research question investigated whether age was a significant moderator for any of the found score–achievement relationships. For all five scores, (NNAT, CogAT Composite, Verbal Battery, Nonverbal Battery, and Quantitative Battery) interaction variables were added into the regression equations for each of the six measured achievement outcome (Broad Reading, Basic Reading, Reading Comprehension, Broad Mathematics, Math Problem Solving, and Math Calculation. Thus, 30 interactions were tested. Out of these 30 interactions, only one was found to be significant.

The relationship between the NNAT Score and Basic Reading achievement was found to be significantly moderated by students' age (when achievement was measured). Age and NNAT score were entered in the first step of the regression analysis. This first model explained 30.9% of the variance in Basic Reading achievement In the second step of the regression analysis, the interaction term (Age x NNAT Score) was introduced. The introduction of this interaction term explained an additional 5.2% of the variance in Basic Reading achievement. This change in R^2 was found to be significant (p =.03). However, given the number of tests, and the overall pattern of non-significant interactions, it is likely that this significant finding is the result of a Type 1 error.

DISCUSSION

Group-administered ability tests provide an efficient, cost-effective method of assessment for identifying students who may benefit from gifted, or advanced, educational programming (Cao, Jung, & Lee, 2017). The purpose of this study was to investigate the relationships between two frequently administered group ability tests, the Naglieri Nonverbal Ability Test (NNAT) and the Cognitive Abilities Test (CogAT), and academic achievement in reading and mathematics.

The major findings of this study are as follows: 1) Generally, students who received full-time gifted programming were observed to obtain significantly higher academic achievement in comparison to students' who receive no full-time programming. These differences were most pronounced in students' math calculation skills. 2) The NNAT score and CogAT Composite score appear to have similar predictive properties in relation to students' reading and mathematics achievement. 3) Utilizing both the NNAT and the CogAT generally provides more information regarding students' academic achievement than either test alone. 4) The CogAT's Verbal Battery was found to be the least related to achievement in reading and mathematics. Whereas, the Quantitative and Nonverbal Battery appear to relate similarly to academic achievement in these domains. 5) Compared to its batteries, the CogAT Composite score is the strongest predictor of academic achievement in reading and mathematics. 6) The Nonverbal Battery of the

CogAT does not provide information above and beyond students' NNAT scores in terms of academic achievement.

1) Generally, students who received full-time gifted programming had significantly higher academic achievement in comparison to students' who receive no full-time programming. These differences were most pronounced in students' math calculation skills.

As hypothesized, group-differences in academic achievement were present between students who did not receive gifted programming and those who did. Students not in gifted programming had significantly lower achievement in the measured outcomes of Broad Reading, Reading Comprehension, Math Problem Solving and Math Calculation in comparison to students who received full-time gifted programming. Further these students had significantly lower Broad Mathematic achievement in comparison to both part-time and full-time gifted students. No group differences in Basic Reading achievement were found between those who did not receive gifted programming and those who did. It is important to note that this study assumes these differences arose due to the different instruction students were receiving and not students ability levels.

The lack of group differences in Basic Reading achievement may be attributed to the fact that gifted programming officially begins in the third grade in the school district that participants were enrolled in. Generally speaking, instruction in third grade focuses less on basic reading skills, as this is grade where education shifts from "learning to read" to "reading to learn." Therefore, the participants likely have had similar instructional experiences in basic reading that are not related to what level of gifted programming they receive.

The hypothesis that group-differences would be most pronounced in Math Calculation skills was supported. The largest difference in academic achievement was observed in Math Calculation skills between non-gifted students and full-time gifted students. On average, students in full-time gifted programming scored 21.23 points higher on the WIAT-III's Numerical Operations subtest than students' not receiving any gifted academic programming. The magnitude of this difference may be explained by the instructional differences between gifted and non-gifted programming in mathematics. Full-time gifted students are provided with accelerated mathematics instruction that results in students' learning material that is two to three grade levels above their non-gifted peers. Therefore, they are exposed different types of math calculation skills before their same-aged peers (such as memorization of math facts, long division, solving for x, etc.). Thus, it is somewhat expected that these students would score higher on measures age-referenced achievement measures, such as the WIAT-III.

The lack of significant differences between part-time gifted students and other groups (with the exception of Broad Mathematic achievement) may be the result of heterogenous instruction received at the part-time level. Part-time gifted programming took the form of gifted programming in one or more areas of academic strength. Participants in the part-time group differed in the number, and area of academic strength and thus the type of instruction received by students in this group varies.

2) The NNAT score and CogAT Composite score appear to have similar predictive properties in relation to students' reading and mathematics achievement.

Both the NNAT and CogAT Composite score were found to be significant predictors of Broad Reading, Basic Reading, Broad Mathematics, Math Problem Solving, and Mathematic Calculation achievement. The hypothesis that the CogAT would be a better predictor of reading and mathematics achievement was unsupported. In fact, the standardized coefficients for each test score were found to be very similar when predicting across the measured academic domains. The similarity in predictive abilities suggest that both measures provide similar information about how students' cognitive abilities may relate to academic performance.

3) Utilizing both the NNAT and the CogAT generally provides more information regarding students' academic achievement than either test alone.

The results of this study indicate that regression models with both the NNAT score and the CogAT Composite score as predictors explain more variance in Basic Reading, Broad Mathematics, and Math Problem Solving achievement in comparison to just one score. In these models, the addition of the CogAT Composite score significantly explains 7-8% more variance than the NNAT score alone. In these models, both the NNAT score and CogAT Composite score retain their significance. Thus, it appears that both scores provide unique, and significant information in regards ability–achievement relationships in these domains.

4) The CogAT's Verbal Battery was found to be the least related to achievement in reading and mathematics. Whereas, the Quantitative and Nonverbal Battery appear to relate similarly to academic achievement in these domains.

The Verbal Battery of the CogAT was found to be the least related to academic achievement, as it was only a significant predictor of one of the six measured outcomes (Math Problem Solving). The Nonverbal and Quantitative batteries both were found to be significant predictors five measured outcomes (Broad Reading, Basic Reading, Broad Mathematics, Math Problem Solving, and Math Calculation).

The hypothesis that the Verbal Battery of the CogAT would explain the most variance in Reading achievement was unsupported. In fact, the Verbal Battery was not found to significantly predict any of the measured reading outcomes (Broad Reading, Basic Reading, or Reading Comprehension). This hypothesis was rooted in the known relationships between G*c* and reading. The known relationships between G*c* and reading often emphasize the importance of general verbal knowledge and vocabulary. While, tasks on the Verbal Battery certainly require a student to have relevant background knowledge and vocabulary, it may be that it performance CogAT's' Verbal battery does not represent differences in these abilities, but rather the differences in how students' *apply* background knowledge and vocabulary in their reasoning. If so, this could account for why the Verbal Battery was not found to predict reading achievement. This is consistent with Lohman's (2012) aim to measure verbal reasoning skills as a sub-factor of a students' Fluid Reasoning (G*f*) abilities, and his belief that "reasoning is best indicated not by knowledge of infrequent or abstruse words, but rather the ability to make precise discriminations in meaning among commonly used words" (p.18).

The hypothesis that the Verbal battery would significantly predict mathematics achievement was partially supported. While, the Verbal Battery was found to be a significant predictor of Broad Mathematics Achievement, it appears that this relationship is rooted in its prediction of Math Problem Solving abilities. The Verbal Battery was not found to be a significant predictor of Numerical Operations. This prediction of Math Problem Solving abilities, rather than Math Calculation skills, makes sense when thinking about the different cognitive processes that contribute to success in both tasks. Many items on the WIAT-III's Math Problem Solving subtest are presented as word problems. Word problems require students to process linguistic information in the form of understanding, forming, and applying verbal concepts when coming up with solutions. These processes are necessary to successfully complete the tasks of the Verbal Battery. Most notably, in the Sentence Completion subtest, students must process linguistic information, often using deductive and inferential reasoning when selecting their answers. In contrast, tasks that measure Math Calculation do so by isolating numerical operations (addition, subtraction, etc.), reducing language-processing demands. Thus, these differences may explain why the Verbal Battery was found to be a significant predictor of Math Problem Solving, and not Math Calculation abilities.

The hypothesis that the Quantitative Battery of the CogAT would predict achievement in mathematics was supported. Results revealed that the Quantitative Battery was a significant predictor of Broad Mathematics Achievement, as well as Math problem Solving and Math Calculation abilities. This hypothesis was rooted in previous research that emphasizes Fluid Reasoning's (G*f*) role in mathematics development. This is not surprising given that CHC theory classifies quantitative reasoning – the use of inductive and deductive reasoning with quantitative concepts – as a narrow ability falling under G*f*. Because mathematics inherently is concerned with number, quantity and space it makes sense that this narrow ability would contribute to students' broad mathematic achievement.

Standardized regression coefficients suggest that the Quantitative Battery was the strongest predictor of Math Problem Solving abilities in comparison to the CogAT's other batteries; and only slightly less predictive of Math Calculation abilities in comparison to the Nonverbal Battery. These predictions may be rooted in the similarity between tasks on the Quantitative battery and measures on the WIAT-III. Some of the items on the WIAT-III's Math Problem Solving subtest are remarkably similar to subtests of the Quantitative Battery. For example, an item on the Math Problem Solving subtests asks students to determine how many beads go on an empty peg to complete a pattern, which is the same format as the Number Series subtest on the Quantitative Batteries. Furthermore, other items on the WIAT-III's Math problem Solving subtest require students to identify numerical weights to shapes in equations, just like the picture-based forms of the Number Analogies subtest on the CogAT's Quantitative Battery. In a similar manner, the Number Puzzles subset of the Quantitative Battery requires students to choose the missing number in math equations, which can also be considered an indirect measure math calculation skills. It is important to note, that this overlap between the measurement of ability and achievement (especially at the narrow-ability level) is common.

The Quantitative Battery was found to significantly predict both Broad and Basic Reading abilities. Duncan and colleagues (2007) found that students' early math skills were better predictors of later reading achievement than early reading skills. Given the overlap between the Quantitative Battery and mathematics achievement measures discussed above, and the fact that most participants took the Quantitative Battery in the second grade, the finding that the Quantitative Battery predicts reading achievement may provide further support for Duncan and colleagues finding that early math skills are important in reading development.

The hypothesis that the CogAT's Nonverbal Battery would significantly predict reading and mathematics achievement was mostly supported. The Nonverbal Battery was found to significantly predict all measured achievement outcomes, with the exception of Reading Comprehension. The Nonverbal Battery's prediction of all three mathematics outcome measures is likely rooted in the measurement of Fluid Reasoning (G*f*) in its' Figure Matrices and Picture Classification subtests. This is consistent with prior research suggesting performance nonverbal matrix tasks relate to mathematic achievement (Fuchs et al., 2006). In contrast, the Nonverbal Battery's prediction of Broad and Basic reading achievement is surprising given that G*f* has only been found to relate to Reading Comprehension abilities at older ages (McGrew & Wendling, 2012).

5) Compared to its batteries, the CogAT Composite score is the strongest predictor of academic achievement in reading and mathematics.

The hypothesis that the CogAT Composite score would not be the best predictor of reading and mathematics achievement was unsupported. The analyses revealed that models predicting reading and academic achievement from CogAT Composite scores explained more variance than those using the CogAT Battery Scores as predictors. Additionally, the CogAT Composite score had higher unstandardized and standardized coefficients in comparison to the Battery score models. Therefore, the CogAT Composite score was actually the best predictor of Reading and Mathematics achievement.

The hypotheses that the CogAT Battery scores would better predict achievement was rooted in CHC theory, which, according to McGrew and Schneider (2012), states that the concept of general intelligence (*g*) is not particularly meaningful in applied contexts. However, the application of CHC theory (i.e. the tendency to ignore *g* in favor of other broad abilities) has recently been criticized. This criticism comes as a result of researchers reanalyzing Carroll's work with the goal of assessing the interpretive and clinical utility of CHC broad abilities in school settings (Benson et al., 2019). Benson and colleagues (2019) found that broad ability factors such as General Comprehension Knowledge (G*c*), Fluid Reasoning (G*f*), and Visual-Spatial (G*v*) were not consistently meaningful in terms of their interpretive and clinical utility in school settings. However, *g*'s factor was consistently meaningful (Benson et al., 2019). This finding does not support overlooking *g*'s role in applied settings such as schools.

In light of these recent criticisms, the finding that the CogAT Composite score best predicts academic achievement is somewhat more expected, as Lohman (2012) states that the composite score of the CogAT can be considered a measure of overall, general intelligence (*g*).

6) The Nonverbal Battery of the CogAT does not provide information above and beyond a students' NNAT score in terms of academic achievement.

The hypothesis that the Nonverbal Battery of the CogAT would not add any predictive value above and beyond that of the NNAT was supported. Introducing the Nonverbal Battery score to the models where the NNAT was found to significantly predict academic achievement failed to significantly explain more variance. Furthermore, when the Nonverbal Battery Score was added to models predicting Basic Reading, Broad Mathematics, Math Problem and Math Calculation achievement, it no longer was a significant predictor of the outcome.

This finding may be attributed to differences in measured broad abilities. Namely, the inclusion of (G*v*) abilities in the CogAT's Nonverbal Battery (Paper Folding). Lakin and Gambrell (2012) found that the paper folding subtest had the largest loading on the Nonverbal Battery's domain specific factor. Thus, the nonverbal battery likely is a better measure of G*v* than it is Fluid Reasoning (G*f*) despite capturing G*f* abilities in its other subtests. In contrast, the NNAT is a measure serves as solely a measure of G*f*. This may explain why the NNAT retains its significance whereas the Nonverbal Battery loses its significance (as the NNAT measuring a purer form of G*f*, and G*f* is a better predictor of achievement). In other words, the inclusion of G*v* in the CogAT's Nonverbal Battery may hinder its predictive ability. This is especially the case when the Nonverbal Battery is introduced into a model that embodies a better predictor of G*f* abilities (in this case the NNAT score).

Limitations

This study is not without limitations. Several limitations exist in regards to the sample. The nature of the recruitment process results in an overall lack of diversity within the participant sample in several areas. First, students who participated in this study generally had highly-educated parents, with professional occupations. Although these characteristics are somewhat representative of the local student population, they are not reflective of a national sample of school-aged children. Most notably, only 4% of participants qualified for free- or reduced- price lunch in their public school. Given the extensive research on the effects of socioeconomic status on academic achievement, the findings of this study are not generalizable from students who differ from the sample obtained. Second, although some participants were reported to speak or understand a language other than English, all parents reported English as their child's most dominant language. Therefore, the findings of this study cannot generalize to students from linguistically diverse backgrounds. Further research should be conducted exploring ability–achievement relationships in linguistically diverse students, such as English-language learners. Lastly, the recruitment process predominately took place via a university-run training clinic that provides group- and individually-administered cognitive assessments to local families. Families coming to the clinic often are seeking these assessments as part of the application process for local gifted education programs. Relatedly, students that come to the clinic typically have above-average to superior cognitive abilities. This is evident in the sample means across group-ability scores that fall at or above the 88th percentile. Therefore, recruited participants do not represent the full-range of cognitive abilities seen in general student populations, and may not generalize to a more general population of students. It is important to keep in mind, however, that given the use of these scores for gifted identification, the relations between high scores and academic achievement are perhaps the most useful to schools when making placement decisions.

Implications

The results of this study provide information to schools regarding how group-administered ability tests relate to student achievement in reading and mathematics. Group ability tests provide schools an efficient, cost-effective way of identifying students who may benefit from gifted, or advanced, educational programming. This study's findings suggest that students' performance on group-ability tests generally relates to their academic achievement in reading and mathematics. This relationship generally provides support for the use of group-administered ability tests as one potential method of identification of students for gifted programming. Especially in school districts that conceptualize giftedness as academic potential, and when gifted programming takes the form of an accelerated academic curriculum. However, it's important to note that at best performance on group-ability tests only accounted for less than half of the variance in mathematics achievement, and less than one-third of the variance in reading achievement. This is consistent with more modern conceptualizations of giftedness which posit that other non-cognitive factors contribute to achievement. Therefore, school districts are not advised to use group-administered ability tests as the sole method of identification.

More specifically, this study also provides schools information regarding two of the most popular group-administered ability tests. First, the CogAT Composite score and the NNAT are very similar with regard to their predictive abilities. Therefore, both are equally justified in their use. Second, findings suggest that administering both tests can provide more information relating to students' achievement in comparison to administering only one. If schools are unable to administer both tests, they may wish to consider other factors, such as administration time, when choosing which ability test best suits their needs. The NNAT takes 30 minutes to administer; whereas the CogAT generally takes 1.5 hours (30 minutes per battery) to administer. The CogAT does not need to be administered in one-sitting.

Schools often opt for Nonverbal measures when identifying students of culturally and linguistically diverse backgrounds for gifted programming. The NNAT is purely nonverbal, whereas the CogAT yields a Nonverbal Battery score that can be used for this purpose. Although both test scores were found to be able to predict achievement independently. The results of this study suggest that the NNAT is a better predictor of academic achievement in comparison to the CogAT's Nonverbal Battery, and that the Nonverbal Battery does not provide more predictive information regarding students' academic achievement above and beyond that of the NNAT. This is likely attributed to the NNAT's better measurement of G*f* abilities in comparison to the NNAT. Therefore, when assessing students from culturally and linguistically diverse backgrounds, schools are advised to pick nonverbal measures, such as the NNAT or the Ravens Progressive Matrices, that measure solely G*f* abilities as they appear to be the most predictive of academic achievement. However, schools should be mindful that group-differences exist on nonverbal measures.

In conclusion, this study provides valuable information to schools who are currently using or thinking about using group-administered ability tests as part of their identification process for gifted academic programming. School districts should consider characteristics of their student population, as well as characteristics of their gifted academic programming when applying this information to their identification practices.

REFERENCES

Ambrose, D., VanTassel-Baska, J., Coleman, L. J., & Cross, C. T. (2010). Unified, insular, firmly policed, or fractured, porous, contested, gifted education? *Journal for the Education of the Gifted, 33,* 453-478.

Araújo, S., Reis, A., Petersson, K. M., & Faísca, L. (2015). Rapid automatized naming and reading performance: A meta-analysis. *Journal of Educational Psychology, 107*(3), 868–883.

Bianco, D. S. (2010). Improving student outcomes: Data-driven instruction and fidelity of implementation in a Response to Intervention (RTI) model. *Teaching Exceptional Children Plus, 6,* 1-13.

Binet, A., & Simon, T. (1916). *The development of intelligence in children* (E. S. Kite, Trans.). Baltimore: Williams & Wilkins.

Borland, J. H. (2005). Gifted education without gifted children: The case for no conception of giftedness. In R. J. Sternberg & J. E. Davidson (Eds.), *Conceptions of Giftedness* (2nd ed., pp. 1– 19). New York: Cambridge University Press.

Bracken, B. A., & McCallum, S. (1998). *Universal Nonverbal Intelligence Test.* Chicago, IL: Riverside Publishing.

Brown, L., Sherbenou, R. J., & Johnsen, S. K. (2010). *Test of nonverbal intelligence* (4th ed.). Austin, TX: Pro-Ed.

Brown-Chidsey, R., Bronaugh, L., & McGraw, K. (2009). It takes a village: Building consensus in school and at home. In *RTI in the Classroom: Guidelines and Recipes for Success* (pp. 6–16). New York, NY: The Guilford Press.

Buysse, V., & Peisner-Feinberg, E.S. (2013). Response to intervention: Conceptual foundations for the early childhood field. In V. Buysse & E.S. Peisner-Feinberg (Eds.), *Handbook of response to intervention in early childhood* (pp. 3-23). Baltimore, MD: Paul H. Brookes.

Caemmerer, J. M., Maddocks, D. L. S., Keith, T. Z. & Reynolds, M. R. (2018). Effects of cognitive abilities on child and youth academic achievement: Evidence from the WISC-V and WIAT-III. *Intelligence, 68*, 6-20.

Cao, T. H., Jung, J. Y., & Lee, J. (2017). Assessment in gifted education: A review of the literature from 2005 to 2016. *Journal of Advanced Academics, 28*(3), 163-203.

Card, D., & Giuliano, L. (2015). *Can universal screening increase the representation of low income and minority students in gifted education?* National Bureau of Economic Research.

Carmines, E. G., & Zeller, R. A. (2011). *Reliability and validity in assessment*. Thousand Oaks: Sage Publications.

Cattell, R. B. (1943). The measurement of adult intelligence. *Psychological Bulletin, 40,* 153-193.

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. Journal of Educational Psychology, 54,1-22.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analysis studies.* New

York: Cambridge University Press.

Chambers, J. A. (1969). A multidimensional theory of creativity. *Psychological Reports, 25*, 779– 799 .

Chapman, P. D. (1988). *Schools as sorters: Lewis M. Terman, applied psychology, and the intelligence testing movement, 1890–1930.* New York, NY: New York University Press.

Choice, P. (2011). *The new RtI: Response to intelligence.* Marion, IL: Pieces of Learning.

Chong, S. L. & Siegel, L. S. (2008). Stability of computational deficits in math learning disability from second through fifth grades. *Developmental Neuropsychology*, 21, 201-218.

Dai, D. Y. & Chen, F. (2013). Three paradigms of gifted education: In search of conceptual clarity in research and practice. *Gifted Child Quarterly, 57*(3), 151-168.

Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligences, 35,* 13-21.

Duncan, G., Dowsett, C., Claessens, A., Magnuson, K., Huston, A., Klebanov, P., … Japel, C. (2007). School Readiness and Later Achievement. *Developmental Psychology*, *43*(6), 1428–1446.

Feldhusen, J. F. (1998). A conception of talent and talent development. In R. C. Friedman & K. B. Rogers (Eds.), *Talent in context: Historical and social perspectives on giftedness* (pp. 193-211). Washington, DC: APA.

Feldhusen, J. D. & Jarwan, F. A. (2000) Identification of gifted and talented youth for educational programs. In K. A. Heller, F. J. Mönks, R. J. Sternberg, & R. F. Subotnik (Eds.), *International handbook of giftedness and talent* (pp. 271–282). Oxford: Elsevier Science.

Feldman, D. H. (2000). The development of creativity. In R. J. Sternberg (Ed.), *Handbook of creativity* (pp. 169-189). Cambridge, UK: Cambridge University Press.

Floyd, R. G., Bergeron, R., Alfonso, V. C. (2006). Cattell–Horn–Carroll ability profiles of poor comprehenders. *Reading and Writing, 19*(5), 427-456.

Floyd, R. G., Evans, J. J., & McGrew, K. S. (2003). Relations between measure of Cattell–Horn–Carroll (CHC) cognitive abilities and mathematics across the school-age years. *Psychology in the Schools, 40*(2), 155-171.

Fuchs, D., & Fuchs, L. S. (2006). Introduction to response to intervention: What, why, and how valid is it. *Reading Research Quarterly*, *41*(1), 93–99.

Fuchs, L. S., Fuchs, D., Compton, D. L., Powell, S. R., Seethaler, P. M., Capizzi, et al. (2006). The cognitive correlates of third-grade skill in arithmetic, algorithm computation and arithmetic word problems. *Journal of Educational Psychology, 98*(1), 29-43.

Flanagan, D. P., & Harrison, P. L. (2012). *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed.). New York: Guilford Press.

Flanagan, D. P., Ortiz, S. O., Alfonso, V. C., & Mascolo, J. T. (2006). The achievement test desk reference, 2nd ed. (ATDR-2). New York: Wiley

Gagné, F. (2005). From gifts to talents: The DMGT as a developmental model. In R. J.

Sternberg & J. E. Davidson (Eds.) *Conceptions of giftedness* (2nd ed., pp. 98-120). Cambridge, UK: Cambridge University Press.

Geary, D. C., (2007). An evolutionary perspective on learning disabilities in mathematics. *Developmental Neuropsychology, 32*(1), 471-519.

Gelman, R., & Butterworth, B. (2005). Number and language: How are they related? Trends in cognitive sciences, learning-disabled students. *Cognition & Instruction, 5*, 223-266.

Giessman, J. A., Gambrell, J. L., & Stebbins, M.S. (2013). Minority performance on the Naglieri Nonverbal Abilities Test, Second Edition, versus the Cognitive Abilities Test, Form 6: One gifted program's experience. *Gifted Child Quarterly*, *57*(2), 101-109.

Gustafsson, J. E. (1984). A unifying model for the structure of intellectual abilities. *Intelligence, 8*, 179-203.

Hebb, D. O. (1942). The effects of early and late brain injury upon test scores, and the nature of normal adult intelligence. *Proceedings of the American Philosophical Society, 85*, 275-292.

Hoover, W., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing: An Interdisciplinary Journal, 2*, 127-160.

Horn, J. L. (1965). *Fluid and crystallized intelligence: A factor analytic study of the structure among primary mental abilities*. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign. (AAT 6507113)

Horn, J. L. & Blankson, A. N. (2012). Foundations for better understanding cognitive abilities. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: theories, tests, and issues* (pp. 99-144). New York: The Guilford Presss.

Jenkins, J. R., Fuchs, L. S., Van den Broek, P., Espin, C., & Deno, S. L. (2003). Sources of individuals differences in reading comprehension and reading fluency. *Journal of Educational Psychology, 95*, 719-729.

Jolly, K. (2018) *The history of gifted education.* New York, NY: Routledge.

Joshi, R. M. & Aaron, P. G. (2000). The component model of reading: Simple view of reading made a little more complex. *Reading Psychology, 21,* 85-97.

Kaufman, S. B., Reynolds, M. R., Liu, X., Kaufman, A. S., & McGrew, K. S. (2012). Are cognitive g and academic achievement g one and the same g? An exploration on the Woodcock–Johnson and Kaufman tests. *Intelligence, 40,* 123-138.

Keith, T. Z., & Reynolds, M. R. (2010). Cattell–Horn–Carroll abilities and cognitive tests: What we've learned from 20 years of research, *Psychology in the Schools, 47*(7), 635-650.

Kintsch, W., & Rawson, K. A. (2005). Comprehension. In M. J. Snowling & C. Hulme (Eds.), The science of reading: A handbook (pp. 209-226) Oxford, England: Blackwell Publishing.

Lakin, J. M., & Gambrell, J. L. (2012). Distinguishing verbal, quantitative, and figural facets of fluid intelligence in young students, *Intelligence, 40,* 560-570.

Lohman, D. F. (2005). The role of nonverbal ability tests in identifying academically gifted students: An aptitude perspective. *Gifted Child Quarterly, 49*(2), 111-138.

Lohman, D. F. (2012). *Cognitive Abilities Test, Form 7: Research and development guide*. Rolling Meadows, IL: Riverside Publishing.

Lohman, D. F., Korb, K. A., & Lakin, J. M. (2008). Identifying academically gifted English-language learners using nonverbal tests: A comparison of the Raven, NNAT, and CogAT. *Gifted Child Quarterly*, *52*(4), 275-296.

Matthews, D. J., & Foster, J. F. (2006). Mystery to mastery: Shifting paradigms in gifted education. *Roeper Review*, *28*, 64-69.

Marland, S. P., Jr. (1972a). *Education of the gifted and talented: Report to the Congress of the United States by the U.S. Commissioner of Education and background papers submitted to the U.S. Office of Education* (2 vols.). Washington, DC: U.S. Government Printing Office.

McClain, M.-C., & Pfeiffer, S. (2012). Identification of gifted students in the United States today: A look at state definitions, policies, and practices. *Journal of Applied School Psychology*, 28, 59–88.

McGill, R. J., Dombrowski, S. C., & Canivez, G. L. (2018). Cognitive profile analysis in school psychology: History, issues, and continued concerns. *Journal of School Psychology, 71*, 108-121.

McGrew, K. S. (1993) The relationship between the WJ-R cognitive clusters and reading achievement across the lifespan. *Journal of Psychoeducation Assessment Monograph Series: Woodcock-Johnson Psycho-Educational Assessment Battery–Revised,* 39-53.

McGrew, K. S., LaForte, E. M., & Schrank, F. A. (2014). Technical Manual. *Woodcock-*

Johnson IV. Rolling Meadows, IL: Riverside

McGrew, K. S. & Wendling, B. J. (2010). Cattell–Horn–Carroll cognitive–achievement relations: What have we learned from the past 20 years of research. *Psychology in the Schools, 47*(7), 651-674.

Naglieri, J. A. (2018). *Naglieri Nonverbal Ability Test Manual* (3rd ed.). Bloomington, MN: Pearson.

Naglieri, J. A., & Ford, D. Y. (2003). Addressing underrepresentation of gifted minority children using the Naglieri Nonverbal Ability Test (NNAT). *Gifted Child Quarterly, 47,* 155-160.

Namkung, J. M., & Fuchs, L. S. (2016) Cognitive predictors of calculations and number line estimation with whole numbers and fractions among at-risk students. *Journal of Educational Psychology, 108*(2), 214.

Nation, K., Clarke, P., & Snowling, M. J. (2002). General cognitive ability in children with reading comprehension difficulties. *British Journal of Educational Psychology, 72,* 549-560.

Osmon, D. C., Smertz, J. M., Braun, M. M., & Plambeck, E. (2006). Processing abilities associated with math skills in adult learning disability. *Journal of Clinical ad Experimental Neuropsychology, 28,* 84-95.

Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading, 11*(4), 357-383.

Pfeiffer, S. I. (2015). *Essentials of gifted assessment.* Hoboken, New Jersey: Wiley & Sons Inc.

Proctor, B. E., Floyd, R. G., & Shaver, R. B. (2005). Cattell–Horn–Carroll broad cognitive ability profiles of low math achievers. *Psychology in the Schools, 42*(1), 1-12.

Reis, S. M., Burns, D. E., & Renzulli, J. S. (1992). *Curriculum compacting: The complete guide to modifying the regular curriculum for high ability students*. Mansfield Center, CT: Creative Learning Press.

Renzulli, J. S. (1978). What makes giftedness? Reexamining a definition. *Phi Delta Kappan, 60,* 180-184.

Renzulli, J. S. (2005). The three-ring definition of giftedness: A developmental model for promoting creative productivity. In R. J. Sternberg & J. E. Davidson (Eds.), *Conceptions of giftedness* (2nd ed., pp.246-280). New York: Cambridge University Press.

Rogers, K. B. (2007). Lessons learned about education the gifted and talented: A synthesis of the research on educational practice. *Gifted Child Quarterly, 51,* 382-396.

Schneider, W. J. & McGrew, K. S. (2012). The Cattell–Horn–Carroll model of intelligence. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: theories, tests, and issues* (pp. 99-144). New York: The Guilford Press.

Seethaler, P. M., Fuchs, L. S., Star, J. R., & Bryant, J. (2011). The cognitive predictors of computational skill with whole versus rational numbers: An exploratory study. *Learning and Individual Differences, 21,* 536-542.

Singley, A. T. M., & Bunge, S. A. (2014). Neurodevelopment of relational reasoning: Implications for mathematical pedagogy. *Trends in Neuroscience and Education, 3*(2), 33-37.

Spearman, C. (1904). “General intelligence,” objectively determined and measured. *American Journal of Psychology, 15*(2), 201-293.

Stein, M. I. (1968). Creativity. In E. Borgalta & W. W. Lambert (Eds.), *Handbook of personality theory and research*. Chicago: Rand McNally.

Sternberg, R. J. (2003). *Wisdom, intelligence, and creativity, synthesized*. New York: Cambridge University Press.

Sternberg, R. J. (2005). The WICS model of giftedness. In R. J. Sternberg & J. E. Davidson (Eds.), *Conceptions of giftedness* (2nd ed., pp. 327–243). New York: Cambridge University Press.

Sternberg, R. J. & Kaufman, S. B. (2018). Theories and conceptions of giftedness. In S. I. Pfeiffer (Eds.), *Handbook of giftedness in children* (pp. 29-47). Cham: Springer.

Subotnik, R. F. (2009). Developmental transitions in giftedness and talent: Adolescence into adulthood. In F. D. Horowitz, R. F. Subotnik, & D. J. Matthews (Eds.), *The development of giftedness and talent across the life span* (pp. 155– 170). Washington, DC: American Psychological Association.

Swanson, H. L., & Jerman, O. (2006). Math disabilities: A selective meta-analysis of the literature. *Review Research, 76*, 249-274.

Tannenbaum, A. J. (1986). Giftedness: A psychosocial approach. In R. J. Sternberg & J.

E. Davidson (Eds.), *Conceptions of giftedness* (pp. 21-252). New York: Cambridge University Press.

Terman, L. M. (1916). *The measurement of intelligence*. Boston: Houghton Mifflin.

Terman, L. M. (1925). Genetic studies of genius: Vol. 1. *Mental and physical traits of a thousand gifted children.* Stanford, CA: Stanford University Press

Thurstone, L. M. (1938). *Primary mental abilities.* Chicago: University of Chicago Press.

VanTassel-Baska, J. (2005). Domain-specific giftedness: Applications in school and life. In R. J. Sternberg & J. E. Davidson (Eds.), *Conceptions of giftedness* (2nd ed., pp. 358–377). New York: Cambridge University Press

Vellutino, F. R., Tunmer, W. E., Jaccard, J. J., & Chen, S. (2007). Components of reading ability: Multivariate evidence for a convergent skills model of reading development. *Scientific Studies of Reading, 11*(1), 3-12.

Walberg, H. J. (1969). A portrait of the artist and scientist as young men. *Exceptional Children, 35,* 5– 12 .

Walberg, H. J. (1971). Varieties of adolescent creativity and the high school environment. *Exceptional Children, 38,* 111– 116 .

Wechsler, D. (2009). *Wechsler Individual Achievement Test (3rd ed.).* San Antonio, TX: Pearson.

Wechsler, D. (2013). *Wechsler Intelligence Scale For Children – Fifth Edition*. Bloomington, MN: Pearson.

Weschler, D., & Naglieri, J. A. (2006). *Weschler Nonverbal Scale of Ability: Technical and interpretive manual.* San Antonio, TX: The Psychological Corporation.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III tests of cognitive abilities*. Itasca, IL: Riverside Publishing.

Ziegler, A., & Raul, T. (2000). Myth and reality: A review of empirical studies on giftedness. *High Ability Studies*, *11*, 113-136.

BIOGRAPHY

Elpiniki Marchesini grew up in Fairfax County, Virginia. She received her BS in Psychology from Virginia Tech (Go Hokies!), where she also minored in Biological Science. After graduating from Virginia Tech, she pursued her master's degree in School Psychology at George Mason University. As a master's student, she became increasingly interested in Developmental Psychology and how its' principles could be applied in school settings. After completing her master's degree, Elpiniki Marchesini joined Dr. Timothy Curby's Development in School Contexts (DISC) Lab as a doctoral student. Her current research interests include: assessment practices used in academic decision-making, whole-school improvement via multi-tiered service delivery models, collaboration between special and general educators, and the benefits of social-emotional teaching on a variety of student outcomes.