AN INVESTIGATION INTO THE RELATIONSHIP BETWEEN COGNITIVE ABILITY, STANDARDIZED ACHIEVEMENT, AND GRADES IN MIDDLE SCHOOL

A Dissertation
Submitted
to the Temple University Graduate Board

In Partial Fulfillment
of the Requirements for the Degree of
Doctor of Philosophy

By
Leslie Blue
May, 2009

ABSTRACT

Title: An Investigation into the Relationship Between Cognitive Ability, Standardized Achievement, and Grades in Middle School

Candidate's Name: Leslie Blue

Degree: Doctor of Philosophy

Temple University, 2009

Doctoral Advisory Committee Chair: Dr. Catherine Fiorello

Today, many school districts are mandating tests to measure student performance and to hold individual schools and school systems accountable for that performance in order to meet the standards set forth in the No Child Left Behind Act (2001) and the Individuals with Disabilities Education Improvement Act (2004). The focus of this study was to examine the relationship among cognitive ability as measured by the Cognitive Abilities Test (CogAT) and measures of achievement, specifically, standardized achievement scores on the New Jersey Assessment of Skills and Knowledge (NJ ASK) and school grades. The current study investigated archival data of 452 seventh grade students enrolled in a large, suburban public school district during the 2007-2008 school year. Scores on the CogAT and NJ ASK were collected from grades 3, 5, and 7. Final grades in the subject areas of Reading, Writing, Math, Social Studies and Science were collected from report cards from the end of seventh grade of the 2007-2008 school year. Pearson correlations found significant relationships between: (1) cognitive ability and standardized achievement scores in grades 3, 5, and 7, (2) third grade cognitive ability and grade seven grades, and (3) third grade standardized achievement scores and grade seven grades. Further, out of the five cluster scores on the grade 3 CogAT and NJ ASK, the NJ ASK Language Arts score was the best predictor of grades in Reading and Writing and the NJ ASK Mathematics score was the best predictor of grades in Math, Science, and Social Studies. Finally, third grade NJ ASK Language Arts, NJ ASK Mathematics and CogAT Verbal scores were the best predictors of special education classification in grade 7, accounting for a combined 22% of the variance. Limitations to the study and implications for future research and practice are discussed.

ACKNOWLEDGMENTS

I would like to acknowledge my husband and my family for their continual support and encouragement during my doctoral work. My husband, Matthew, has offered me support, love, and wisdom. He has also provided occasional intimidation, which I now understand to be in both of our best interests. My parents, Ben and Carol, have offered love, advice and support through this journey. My smart, loving and intuitive sister, Allyson, has served as an exceptional role model worthy of much admiration and emulation over the years. Her precious children, Ben and Kyle, have provided me with real life lessons in child development, education and psychology. They have taught me that, while we try to teach our children all about life, our children teach us what life is all about. Each of these people has made me a better, more empathetic psychologist. My accomplishments would mean very little without each of these individuals to share them with.

I would also like to thank the school psychology faculty for providing an ideal graduate school experience, especially Dr. Catherine Fiorello for her guidance and mentorship during this project. I thank Dr. Joseph DuCette for broadening my understanding of statistical analyses and offering patient guidance. Finally, I thank Dr. Ken Thurman and Dr. Frank Farley for serving on my committee and for generously lending their expertise, time and support.

This is a great opportunity to express my respect and thanks to my cohort for working together as we pursued the same goal in different ways.

Finally, I would like to thank the members of the school district that made this study possible. They recognized the importance of self-reflection and the beneficial role that data analyses can play in improving public education.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Context

This study investigated the relationship between cognitive ability and measures of achievement, specifically, standardized achievement scores on the New Jersey Assessment of Skills and Knowledge (NJ ASK) and school grades.

Measuring what and how well students learn is an important building block in the process of strengthening and improving our nation's schools. Tests, along with student grades and teacher evaluations, can provide critical measures of students' skills, knowledge, and abilities. When tests are developed and used appropriately, they are among the most sound and objective knowledge and performance measures available (http://www.apa.org/pubinfor/testing.html).

The No Child Left Behind Act of 2001 (NCLB) is the latest federal legislation that enacts theories of standards-based education reform, which is based on the belief that setting high expectations and establishing measurable goals can improve individual outcomes in education. It aims to improve the performance of United States primary and secondary schools by increasing the standards of accountability for states, school districts, and schools. The Act requires states to develop assessments of reading and math to be given annually to all students in grades 3 – 8 and once in high school. The assessments must: be based on state content and performance standards, measure higher order thinking, provide useful diagnostic information, and be valid and reliable. NCLB mandates states to have 100% of their students proficient by 2014, and it requires that the same standard of performance be applied to all groups of students, including ethnicity, income, disability and English language proficiency.

The effectiveness and desirability of NCLB's measures are hotly debated. A primary criticism asserts that NCLB could reduce effective instruction and student learning because it may cause states to lower achievement goals and motivate teachers to teach to the test. A primary supportive claim asserts that systematic testing provides data that shed light on which schools are not teaching basic skills effectively, so that interventions can be made to improve outcomes for all students while reducing the achievement gap for disadvantaged and disabled students (http://www.pbs.org/wgbh/pages/frontline/shows/schools/nochild/bush.html).

The NCLB Act, along with the Individuals with Disabilities Education Improvement Act of 2004 (IDEA 2004), has important implications for children with disabilities. IDEA 2004 is the newest reauthorization of the federal special education law pertaining to the provision of special education services to students with disabilities. In IDEA 2004, the language about participating in state and district assessments was changed to:

> All children with disabilities are included in all general State and districtwide assessment programs…with appropriate accommodations and alternate assessments, where necessary and as indicated in their individualized education programs. (Section 1412(c)(16)(A)).

Today, many school districts are mandating tests to measure student performance and to hold individual schools and school systems accountable for that performance in order to meet the standards set forth in NCLB and IDEA. The National Center for Fair and Open Testing estimates that America's public schools administer more than 100 million standardized exams each year, including intelligence, achievement, screening and readiness tests (http://www.fairtest.org/testing-explosion-0).

Public officials and educational administrators are increasingly calling for the use of tests to make high-stakes decisions, such as whether a student will move on to the next grade level or receive a diploma. Currently, 17 states require students to pass a test to graduate and 7 more are planning such tests (http://www.fairtest.org/dangerous-consequences-highstakes-standardized-tes). Critics of high-stakes testing note that school officials using such tests must ensure that students are tested on a curriculum they have had a fair opportunity to learn, so that certain subgroups of students, such as racial and ethnic minority students or students with a disability or limited English proficiency, are not systematically excluded or disadvantaged by the tests or the test taking conditions. Further, high-stakes decisions should not be made on the basis of a single test score, as a single test can only provide a snapshot of student achievement and may not accurately reflect an entire year's worth of student progress and achievement (http://www.apa.org/pubinfor/testing.html).

A potential problem with the current increased emphasis on testing is not necessarily the test, per se, but the instances when tests have unintended and potentially negative consequences for individual students, groups of students, or the education system more broadly. It remains critical to remember that, in many instances, without tests, low-performing students and schools could remain invisible and therefore not get the extra resources or remedial help that they need.

The Standards for Educational and Psychological Testing, created by the American Psychological Association, the American Educational Research Association, and the National Council on Measurement in Education (1999), present a number of principles that are designed to promote fairness in testing and avoid unintended consequences. Test developers must ensure that certain groups of students are not disadvantaged by a test, and test users must guard against allowing the testing process, or the need for students to pass a certain test, to overwhelm the rest of a student's mastery of a wide curriculum. Furthermore, remedial programs should be in place for students who score poorly on such tests.

In light of the calls to improve educational outcomes by measuring student and school performance on standards-based tests, as well as the inclusion of all children with disabilities in all state and districtwide assessment programs, the importance of promoting fairness in testing in order to avoid unintended consequences is paramount. Any decision about a student's continued education, such as retention, tracking, or graduation, should not be based on the results of a single test, but should include other relevant and valid information.

Purpose of Current Study

The focus of this study was to examine the relationship between cognitive ability as measured by the Cognitive Abilities Test (CogAT) and measures of achievement, specifically, standardized achievement scores on NJ ASK and school grades. The current study investigated archival data of 452 seventh grade students enrolled in a large, suburban public school district during the 2007-2008 school year. Scores on the CogAT and NJ ASK were collected from grades 3, 5, and 7. Final grades in the subject areas of reading, writing, math, social studies and science were collected from report cards from the end of seventh grade of the 2007-2008 school year.

Research Questions and Hypotheses

This study addresses four research questions, as follows:

Question 1: What is the relationship between cognitive ability as measured by the CogAT and academic achievement as measured by NJ ASK test scores in grades 3, 5, and 7?

Hypotheses 1A: It was hypothesized that differential validity would be observed for the CogAT Verbal score, which would show the highest correlations with the Language Arts Literacy cluster score on the NJ ASK achievement test at each grade level.

Hypotheses 1B: It was hypothesized that differential validity would be observed for the CogAT Quantitative score, which would show the highest correlations with the Mathematics cluster score on the NJ ASK achievement test at each grade level.

Hypotheses 1C: It was hypothesized that the CogAT Nonverbal battery score would show the lowest correlations with both the Language Arts Literacy and Mathematics cluster scores on the NJ ASK achievement test at each grade level.

Question 2: What is the relationship between cognitive ability as measured by the CogAT in grade 3 and grade 7 grades?

Question 2A: What is the relationship between academic achievement as measured by NJ ASK test scores in grade 3 and grade 7 grades?

Hypotheses 2A: It was hypothesized that differential validity would be observed for the CogAT Verbal score, which would show the highest correlations with grades in the subjects of reading, writing and social studies. It was further hypothesized that differential validity would be observed for the NJ ASK Language Arts Literacy cluster score, which would also show the highest correlations with grades in the subjects of reading, writing and social studies.

Hypotheses 2B: It was hypothesized that differential validity would be observed for the CogAT Quantitative score, which would show the highest correlations with grades in the subjects of mathematics and science. Further, it was hypothesized that differential validity would be observed for the NJ ASK Mathematics cluster score, which would also show the highest correlations with grades in the subjects of mathematics and science.

Hypotheses 2C: It was hypothesized that the CogAT Nonverbal battery score would show the lowest correlations with grades in all subjects (reading, writing, math, science and social studies).

Question 3: Which of the five cluster scores on the grade 3 CogAT and NJ ASK (CogAT Verbal, CogAT Quantitative, CogAT Nonverbal, NJ ASK Language Arts and NJ ASK Mathematics) best predicts achievement in each of the academic subject areas investigated?

Hypotheses 3A: It was hypothesized that the NJ ASK Language Arts Literacy cluster score would be the best predictor of grades in the subjects of reading, writing and social studies.

Hypotheses 3B: It was hypothesized that the NJ ASK Mathematics cluster score would be the best predictor of grades in the subjects of math and science.

Question 4: Can cognitive ability as measured by the CogAT and academic achievement as measured by NJ ASK test scores in grade 3 predict special education classification in grade 7?

Hypotheses 4A: It was hypothesized that CogAT and NJ ASK scores would explain a significant proportion of the variance in predicting special education classification.

Significance of Study

Intelligence, as measured by an intelligence quotient and through Cattell-Horn-Carroll (CHC) Theory, via individually-administered cognitive tests, has been found to correlate with a student's performance on individually-administered standardized achievement tests (e.g., Benson, 2007; Evans, Floyd, McGrew, & Leforgee, 2002; Felton & Pepper, 1995; Flanagan, 2000; Floyd, Evans, & McGrew, 2003; Floyd, Keith, Taub, & McGrew, 2007; Hale, Fiorello, Kavanagh, Hoeppner, & Gaither, 2001; Keith, 1999; McBride-Chang, 1995; McGrew & Flanagan, 1998; McGrew, Flanagan, Keith, & Vanderwood, 1997; Proctor & Shaver, 2005; Vanderwood, McGrew, Flanagan, & Keith, 2002; Wagner & Torgesen, 1987; Williams, McCallum, & Reed, 1996). Since group-administered cognitive and achievement tests, such as the CogAT and the NJ ASK, are regularly used in schools to measure accountability and drive instruction and curriculum, this study investigated the relationship between cognitive ability as measured by the CogAT and academic achievement as measured by NJ ASK test scores.

While there is an increasing body of research linking cognitive abilities with standardized achievement test results, there is little research linking cognitive strengths and weaknesses with grades in particular subject areas. For example, previous research shows a relationship between crystallized ability (G*c*) and reading achievement as measured on standardized achievement tests (Felton & Pepper, 1995; McBride-Chang, 1995; Wagner & Torgesen, 1987). However, there is little research that indicates that above average ability in G*c* correlates with higher grades in reading, language arts or English classes. McLaughlin (2005) studied the relationships between specific cognitive factors within the CHC Theory and grades as markers of achievement in specific subjects in a sample of high school and college students. Results did not demonstrate significant correlations among CHC factors of cognitive abilities and high school grades or college GPA. The second research question in the current study will investigate whether there are relationships between cognitive ability as measured by the CogAT and grades in school. Further research in this area could expand the interpretability of cognitive strengths and weaknesses in relation to academic achievement in school. Also, interpreting cognitive strengths and weakness in relation to grades would provide relevant and valid information to assist in making decisions about a student's education by relying on multiple sources of data.

The primary use of CogAT scores is to provide an alternative measure of cognitive development and to identify students whose predicted levels of achievement differ markedly from their observed levels of achievements (Lohman & Hagen, 2001). Since school psychologists are faced with large numbers of children referred for special education evaluations resulting in increasing caseloads, an attempt to identify methods that would appropriately reduce the number of children referred for special education is of great interest. Reducing the number of referrals through the Intervention and Referral System (I&RS) and/or Pupil Assistance Committees (PAC) will allow practitioners to provide more thorough and comprehensive assessments for those students who truly need them. While scores on group-administered cognitive and achievement tests are available to pre-referral committees, correlating this information to general and special education students' classroom performance is difficult, without a context to interpret such information. Therefore, the third and fourth research questions will be a preliminary investigation into the relationship between cognitive ability as measured by the CogAT, academic achievement as measured by NJ ASK test scores and special education classification. It is the investigator's hope that, by attempting to understand this relationship, group-administered cognitive and achievement scores can be added to the multiple data sources available to intervention committees when determining appropriateness of referral for a more comprehensive evaluation in order to determine special education eligibility.

Definitions

Auditory Processing (Ga). Measures a person's perception, analysis, and synthesis of patterns among auditory stimuli as well as the discrimination of subtle differences between pattern of sounds (Flanagan & Ortiz, 2002), allowing individuals to learn early reading skills through the process of phonetic decoding (Flanagan & Ortiz, 2002; McGrew & Flanagan, 1998).

Cattell-Horn-Carroll (CHC) Theory. Is a contemporary theory for understanding/organizing intelligence and is composed of three levels or strata: Stratum III, *g* or global IQ score; Stratum II, broad ability (G*f*, G*lr*, G*a*, G*v*, G*sm*, G*a*, and G*c* – defined in this section); and Stratum I, narrow ability areas, or specific abilities subsumed by broad abilities. For example, the broad ability, fluid reasoning (G*f*), incorporates the narrow ability areas of deductive and inductive reasoning.

Cognitive Abilities Test (CogAT). A group-administered test designed to measure general abstract reasoning abilities and a student's capacity to apply these abilities to verbal, quantitative, and nonverbal cognitive tasks (Lohman & Hagan, 2001). In the present study, the overall composite score will be examined, along with the verbal, quantitative, and nonverbal reasoning scores.

Cognitive Strength. Refers to relative strengths within an individual's cognitive profile (a score above the individual's overall mean cognitive ability) in relation to the other ability areas assessed and also refers to adequate or above average cognitive ability. Adequate ability is described as a standardized score within the average range (90-110; mean = 100, standard deviation = 15) (Mather & Wendling, 2005; Naglieri, 2005).

Cognitive Weakness. Refers to relative weaknesses and weaknesses based on standardized scores. A relative weakness refers to an ability area that is lower (a score below the individual's overall mean cognitive ability) in relation to the other ability areas assessed. Cognitive weakness also refers to standardized scores that are less than 85 range (mean = 100, standard deviation = 15) (Mather & Wendling, 2005; Naglieri, 2005).

Cross Battery Assessment. The method for bridging theory with practice in order to identify cognitive ability strengths and weaknesses across the broad as well as narrow ability areas. Identifying cognitive strengths and weaknesses is then used to drive intervention selection and to provide a starting point for establishing links between specific cognitive abilities and interventions (Hale & Fiorello, 2004; Mather & Jaffe, 2002; Mather & Kaufman, 2006; Mather & Wendling, 2005; McGrew & Flanagan, 1998).

Crystallized Intelligence/Knowledge (Gc). Crystallized ability refers to information learned through previous experience. Crystallized ability, an area of acquired knowledge, includes one's cultural knowledge obtained through exposure and one's ability to communicate this knowledge. Crystallized ability also includes the capacity to reason with previously learned procedures. This reasoning depends on previously learned information either encountered through formal schooling or in interactions with the environment (Flanagan & Ortiz, 2002; McGrew & Flanagan, 1998).

Fluid Intelligence (Gf). Is the ability to solve problems using unfamiliar information or novel procedures without any explicit instruction. This ability depends very little on cultural knowledge or previous schooling. Fluid reasoning encompasses the ability to use reasoning and problem-solving skills when encountering new information (Flanagan & Ortiz, 2002; McGrew & Flanagan, 1998).

g. Spearman (1904, 1927) defined *g* as a person's intelligence. It is the psychometric component that accounts for most of the variance in performance across a wide variety of cognitive tasks.

Intervention and Referral System (I&RS). The I&RS team is one of many resources used by schools in the state of New Jersey to intervene with student problems, prior to Child Study Team involvement. The I&RS team process is an ongoing, collaborative school effort between district personnel and parents to intervene when a student has been identified as making minimal academic and or emotional progress in the regular education setting. The team develops and monitors interventions to assist the student and continues to identify and evaluate problems, solutions, and progress within the student's academic setting.

IQ score. A single, full scale measure of intelligence.

Long-Term Storage and Retrieval (Glr). Measures a person's ability to hold information for a long period of time as well as the person's ability to smoothly, efficiently, and quickly retrieve that information (McGrew & Flanagan, 1998).

New Jersey Assessment of Skills and Knowledge (NJ ASK). A standardized achievement test designed to give an early indication of the progress students make in mastering the knowledge and skills described in the New Jersey Core Curriculum Content Standards (CCCS) (New Jersey Department of Education, 2004; New Jersey Department of Education, 2006). The current study will examine the two cluster scores in the content areas of Language Arts Literacy and Mathematics.

Processing Speed (Gs). Is the ability to quickly and automatically perform cognitive tasks while maintaining focused attention (Flanagan & Ortiz, 2002; McGrew & Flanagan, 1998).

Short-Term Memory (Gsm). Is the ability to hold information in immediate awareness and then use it within a few seconds (Flanagan & Ortiz, 2002; McGrew & Flanagan, 1998).

Theory of Fluid and Crystallized Intelligence (Gf – Gc Theory). G*f*-G*c* theory was first postulated by Cattell (1941, 1957) to consist of two major types of cognitive abilities – fluid reasoning (G*f*) and crystallized ability (G*c*) (both defined above). Horn (1968) expanded the G*f*-G*c* model to include nine additional cognitive abilities in the domains of visual processing (G*v*), short-term memory (G*sm*), long-term storage and retrieval (G*lr*), speed of processing (G*s*), auditory processing ability (G*a*), quantitative ability or knowledge (G*q*) and facility with reading and writing (G*rw*). Historically, this body of research has been delineated G*f*-G*c* theory. The present study examines only two of the nine broad abilities – G*f* and G*c*.

Visual Processing (Gv). Is the ability to analyze and synthesize visual information. This includes the perception and manipulation of visual shapes and patterns, such as creating three-dimensional designs from pictures or visualizing what a particular pattern would look like if manipulated in some way (Flanagan & Ortiz, 2002; McGrew & Flanagan, 1998).

WJ-R. The Woodcock-Johnson Psychoeducational Battery – Revised, is an empirically supported measure of several constructs within the Cattell-Horn-Carroll (CHC) theory of cognitive abilities.

WJ-III. The Woodcock-Johnson Tests of Cognitive Abilities, Third Edition is a standardized intelligence battery for individuals aged 2 through 90 years old. WJ-III subtests were created in alignment with the CHC Theory of Cognitive Abilities.

CHAPTER 2

REVIEW OF THE LITERATURE

Intelligence as a psychological construct has been examined critically since the early days of testing. It has been defined in many ways, resulting in various forms of measurement. The historically dominant theories of intelligence will be reviewed, including a discussion of intelligence testing and its measurement utility. Academic achievement will then be defined and a taxonomy for understanding specific cognitive abilities and academic achievement will be addressed. This section will end by reviewing the research literature regarding the relationship between cognitive abilities and achievement. The current study made use of one of the more recent definitions of the construct of intelligence, the CHC Theory of Cognitive Abilities, by investigating the measurement of two of these abilities, fluid reasoning (*Gf*) and crystallized ability (G*c*), and their relationship to achievement as defined by the American school system (specifically, NJ ASK scores and school grades).

Definitions of Intelligence

Considerable variability exists on all human traits. Accordingly, “individuals differ from one another in their ability to understand complex ideas, to adapt effectively to the environment, to learn from experience, to engage in various forms of reasoning, to overcome obstacles by taking thought” (Neisser et al., 1996, p. 77). The construct of intelligence has been offered to explain and clarify the complex set of phenomena that account for individual differences in cognitive functioning.

Attempts to define the construct of intelligence and to explain individual differences in intellectual functioning have spanned decades (Carroll, 1993; Gustafsson,

& Undheim, 1996; Kamphaus, 1993; Thorndike, 1997; Thorndike & Lohman, 1990; Sattler, 1988) and have been characterized by much variability. Prominent in definitions are attributes such as adaptation to the environment, basic mental processes, and higher-order thinking (e.g., reasoning, problem solving, and decision making) (Sattler, 2001). It has been described as an innate global capacity (Jensen, 1980; Spearman, 1927), the ability to carry on "abstract thinking" (Terman, 1921), a collection of faculties: judgment, practical sense, initiative and the abilities to adapt to circumstances (Binet & Simon, 1916), "the aggregate or global capacity of the individual to act purposefully, to think rationally and to deal effectively with his environment" (Wechsler, 1958, p. 7), a compilation of individual, yet interrelated cognitive abilities (Carroll, 1993; Cattell, 1967, 1971; Horn, 1985), a developmental process through which people learn to make sense of the world (Piaget, 1952), and a complex set of skills and abilities or multiple intelligences (Gardner, 1983; Sternberg, 1985). While there are many ways to define intelligence, the confusion concerning different definitions is linked to the fact that intelligence is an attribute, not an entity, and that it reflects the summation of the learning experiences of the individual (Wesman, 1968).

Historical Background

In the late 19th century, two issues were critical for the introduction of measures of intelligence – universal compulsory education and the growing sentiment that psychological mental processes could be objectively measured (Thorndike, 1997). Together, these two forces served as the impetus for the development and extensive use of psychometric assessment tools for the measurement of intellectual functioning in school-aged children.

Two pivotal forces associated with the late 19^{th} century compulsory education movement included: (1) the changing of societal attitudes toward youth and (2) massive immigration. One of the preeminent forces associated with the former was the support of the Jeffersonian ideal of the 'educated electorate' (Thorndike, 1997). Jefferson believed that only an educated electorate could make sound self-governing decisions and he was a proponent of universal education. Another tenet of this shifting society status of youth was a growing sentiment that children represented a vulnerable group in society, whose safety should be protected by public policies, resulting in child labor laws. Such restrictions on children working in industrialized centers resulted in extensive free time, creating a need to maintain social order in children. One solution was compulsory education.

Compulsory education was also seen as a way to make Americans out of the large number of immigrants coming into the country each year. Massive immigration resulted in an unprecedented number of diverse students enrolled in American schools. However, the existing curricula and academic standards were not designed to meet the needs of such diversity, which resulted in a high failure rate. This created a desire to preserve resources and develop special schools, particularly for students perceived to be incapable of learning. Intelligence testing, as a means of ascertaining future academic performance, grew out of this need.

The widely perceived needs of the schools to deal with the problems created by universal compulsory education combined with an intellectual climate of naïve optimism about the measurability of mental processes produced an atmosphere in which measures of intellectual ability were almost certain to be developed (Thorndike, 1997). Of course, once such measures were available, it did not take long for them to be widely adopted. The work of psychologists in the second half of the 19th century, including Gustav Fechner, Wilhelm Wundt, and Hermann Ebbinghaus made it possible to quantify various psychological characteristics. In keeping with this trend, the late 19th and early 20th century saw theorists moving into the domains of test construction and corroboration, including human ability and its measurement.

Intelligence Theories and Testing

Intelligence is grounded in the work of Francis Galton in the late 19th century. Galton is considered to be the father of the study of individual differences and the father of the testing movement (Sattler, 2001). He was convinced that the differences in intelligence among individuals were largely hereditary and that a number of hereditary physical traits or abilities were adequate measures of intelligence. Thus, Galton suggested reaction time as a feasible approach and pursued various sensori-motor measurements. He examined differences in individual characteristics and the relationship of those differences to other traits and abilities. Galton’s criterion of mental ability was based on the social and professional life of illustrious men, including judges, famous military commanders, scientists, poets, painters and musicians. The tests he developed generally proved to be invalid and therefore limited his work on the measurement of intelligence (Sattler, 2001). Despite this fact, Galton’s contributions were valuable to the field of psychology, as he moved the assessment of mental ability out of the field of abstract philosophy and demonstrated that mental ability could be studied experimentally and practically. Within Galton’s hereditarian mindset led the inevitable value-laden categorization or ranking of populations based on measurable traits and abilities.

James Keen Cattell, who, in 1890, coined the term "mental test," (Thorndike, 1997), brought Galton's ideas to the United States and proposed a widespread program of mental testing to provide a standard metric for the assessment of human intellectual ability. He described over 50 different measures that primarily assessed sensory and motor abilities, which differed little from those designed by Galton. Cattell set out to prove that tests measured intelligence by showing a relationship between the test results and students' grades. It was Galton's concept of correlation that invalidated Cattell's own method of intelligence testing. However, Galton, and subsequently, Cattell's view of mental testing came to dominate American psychology, and intelligence testing became the means by which their hereditarian views influenced the schooling offered to students, the assignments given to men in the military services, and the immigration policies of the nation (Hunt, 1993).

In France at the end of the 19th century, Alfred Binet, Victor Henri and Theodore Simon developed methods for the study of mental functions. Conceiving of intelligence not as Galton had, in terms of sensory and motor abilities, but as a combination of cognitive abilities, they believed that intelligence was best understood as a collection of various higher-order mental abilities that might be only loosely related to one another (Binet & Henri, 1916). Binet also differed from Galton on the hereditary question, arguing that intelligence was nurtured through interaction with the environment and that an important function of schooling was to increase intelligence. As such, the main purpose of Binet and Simon's test – developed at the request of the French Ministry of Education in response to mandatory universal education of children – was to identify children who were not profiting as much as they should from their schooling so that they might be given special attention. Unlike previous attempts at developing intelligence tests, Binet’s 1905 scale reflected an acknowledgment of age-based cognitive development and became the prototype for subsequent scales for the assessment of mental ability. Defining intelligence in terms of age, and assembling a set of cognitive tasks that measured the mental age of a child, Binet replaced Galton’s anthropometric testing and became the foundation of the intelligence-testing movement (Hunt, 1993).

After the introduction of the Binet-Simon Scales, the testing movement flourished in the United States. Henry Goddard traveled to France, learned about the 1908 Binet-Simon scale, and immediately translated it into English. He was the first to use the Binet-Simon scale for mass testing in order to locate below-normal children and shunt them into special classes. Goddard’s view of intelligence was different from that of Binet, who conceptualized it as a shifting complex of interrelated functions. Instead, Goddard believed that intelligence consisted of a single underlying function. “Further, he believed that this unitary function was largely determined by heredity, a view much at variance with Binet’s optimistic proposals for mental ability” (Tuddenham, 1962, p. 462). Goddard took a position far more severe than Galton and made a consequential social application of the Binet-Simon scale to advocating eugenic sterilizations and deportations of immigrants in the early 1900s.

Despite the acceptance of Goddard’s translation of the Binet-Simon scale, Lewis M. Terman saw certain flaws in it and felt that he could correct them and make the scale more accurate by adapting some items, adding other items, establishing new age norms, and extending the upper age limit. He disseminated Galton's theories of natural ability by defining mental ability and genius in terms of scores on the Stanford-Binet intelligence test. Terman adopted Louis William Stern's concept of mental quotient, which is found by dividing mental age by chronological age. Terman and his associates renamed this ratio the intelligence quotient (IQ) when they produced the 1916 Stanford-Binet Test. The IQ was a score meant to quantify intellectual functioning to allow comparison among individuals. The 1916 Stanford-Binet test became the standard test of measuring intelligence and remained so for over two decades. Although the IQ has become an extremely useful means for classifying persons, Wolf (1969) noted that it is questionable whether Binet "would have accepted even Terman's elaborate standardization as a valid basis for calculating IQ's" (p. 236).

At the same time that Terman was completing the Stanford revision, Robert Yerkes was working on the forerunner of the major competitor of the Stanford-Binet. Yerkes proposed to abandon the age placement of items entirely and use what he called a point scale. In the Yerkes Point Scale (Yerkes, Bridges & Hardwick, 1915), items were grouped by type and ordered by difficulty. The examinee was given a score on each type of item based on the number of correct answers given, with bonus points for exceptionally quick or insightful answers. The score could then be converted into an overall mental level. This is the same format that David Wechsler adopted for the Wechsler-Bellevue (Wechsler, 1939) and its successors.

Like Yerkes, David Wechsler was interested in developing a point scale. Wechsler's original purpose was to design an intelligence test for adults, relying heavily on the work of others as he borrowed many of his items, subtests, point scales and scoring systems. Wechsler felt that the Binet scales were too verbally loaded for use with adults, so he designed an instrument with subtests to measure both verbal and nonverbal abilities. Wechsler's search for subtests was guided by a focus on the global nature of intelligence, since he considered intelligence to be a part of the larger construct of personality. The Wechsler scales were designed to consider factors contributing to the effective intelligence of the individual. He made no attempt to design a series of subtests to measure "primary abilities" or to order the subtests into a hierarchy of relative importance. As such, the overall IQ obtained from the Wechsler scale represented an index of general mental ability. He adopted a mean score of 100 since the Stanford-Binet metric had become universally accepted. He later modified his own adult test to produce a version for children by designing original material for the subtests, although in some cases items differed only slightly from those appearing in the other intelligence tests.

Several reasons explain the popularity of the Wechsler scale: first, adult tests of intelligence were in short supply; second, Wechsler provided a single assessment battery that integrated Verbal and Performance tests; third, he provided a strong normative sample; and fourth, Wechsler emphasized psychometric rigor and introduced the deviation IQ composite score (Wasserman & Tulsky, 2005). The IQ composite score (same age group comparison) as a field standard has continued to dominate mainstream assessment. However, the contributions of contemporary researchers (Hale & Fiorello, 2004; Mather & Wendling, 2005; McGrew and Flanagan, 1998), who emphasize processing abilities, calls into question the utility of distilling an individual's range of cognitive strengths and weakness into a global score (IQ).

Factor Analytic Theories of Intelligence

With the introduction of the Binet-Simon Scale, intelligence testing became a popular assessment technique through the United States. The widespread use of intelligence testing was given further impetus by statistical evidence that the tests measured not just a series of separate mental aptitudes but also an innate core of mental ability or "general intelligence." Until this period in history, the approaches to intelligence had been very pragmatic – tests were developed for particular needs. However, another approach to understanding intelligence involved analyzing data that were already collected, termed factor analysis. Although factor analysts may disagree about how intelligence is organized – whether intelligence is a general unitary function or a composite of several independent abilities – many accept the theory of general intelligence, while still maintaining that intelligent behavior is multidimensional.

Charles Spearman was one of the early proponents of a factor analytic approach to intelligence (Sattler, 2001). He argued that, as a rule, people who do well on specific intelligence tests also do well on a variety of intellectual tasks. With this reasoning, Spearman showed that many mental abilities are correlated. He proposed a two-factor theory of intellect that held that performance on any intelligence measure was composed of two parts: one part due to the individual's level on the trait of general intelligence, which he called *g*, and the other on an ability specific to the particular test, which he called *s*. The correlation between any two tests was attributed to the presence of *g* in both (Thorndike, 1997).

Spearman became increasingly interested in the specific factors and together with Karl Holziner developed a 'bi-factor' model. Carroll (1993) predicted that Spearman would have converged on a multiple abilities model similar to others had he lived beyond 1945.

Using the new method of multiple factor analysis, Thurstone (1938) proposed that there were a small number of “primary mental abilities.” While he accepted Spearman’s hypothesis of a general factor, he disputed its importance. He argued that intelligence is better described and measured by considering distinct primary mental abilities, rather than a single factor g, which does not provide specific information about specific intelligence. Thurstone’s tests have largely dropped out of use because the hope that they would be able to more accurately predict performance than general intelligence was not fulfilled. However, the similarities between his primary mental abilities and CHC theory’s broad cognitive abilities are notable.

The most prominent multifactor theorist in the United States is J. P. Guilford (Sattler, 2001). He parted company from the majority of factorial theorists by refusing to acknowledge the existence of any general factor at all. Instead, he developed the three-dimensional structure of intellect model, stating that intelligence comprises 180 elementary abilities, made up of a combination of three dimensions. Guilford proposed that each combination of a specific operation, a specific type of content, and a specific type of product defines a unique type of intelligence.

In Philip Vernon’s hierarchical theory of intelligence, he suggested that intelligence can be described as composed of abilities at varying levels of generality. At the highest level is *g*, or general ability, as defined by Spearman. At the next level are major group factors. At the lowest level at the bottom of the hierarchy are specific factors again of the kind identified by Spearman. Factors low in the hierarchy refer to narrow ranges of behavior, whiles those high in the hierarchy refer to a wide variety of behaviors. Vernon believed that we must consider a general group factor in any attempt to understand or measure intelligence. This belief has substantial support across numerous studies, as indicated by positive intercorrelations among cognitive tests administered to representative populations (Sattler, 2001).

Howard Gardner has argued against the concept of a general intelligence that cuts across all mental tasks and supports the Thurstonian notion of separate intelligences that may vary independently of one another. He posits several relatively autonomous intellectual competences, or multiple intelligences. He has identified eight competencies and two tentative competencies, but allow that more may be discovered. The competencies include linguistic, musical, logical-mathematical, spatial, bodily-kinesthetic, intrapersonal, interpersonal, naturalist, spiritual, and existential intelligence (Sattler, 2001). The competencies can be viewed as building blocks out of which thought and action develop. The components constitute the basis of human symbol-using capacities, and they interact to produce a diverse mixture of human talents that we can employ to achieve societal ends. Some psychologists contend that Gardner has stretched the meaning of intelligence too far and that latter set of abilities should be called talents rather than intelligences (Gray, 1994). A more damaging criticism of Gardner's theory is that relatively little statistical or other objective evidence exists that the intelligences he proposes are each functionally unitary and independent of each other (Brody, 1992). Still, Gardner's work represents an interesting, pioneering attempt to make some sense of the remarkably diverse and unsystematic case-history literature.

Robert J. Sternberg's triarchic theory of intelligence divides human intelligence into three dimensions: componential, experiential and contextual. The componential dimension relates intelligence to the internal mental mechanisms of the individual, and

Sternberg refers to these mental mechanisms as "information-processing components." The experiential dimension of Sternberg's theory relates intelligence to both the external and internal worlds of the individual, and it specifies the point at which intelligence is most critically involved in an individual's ability to cope with tasks. He argues that intelligence deals with novelty and with the automatization of mental processes. The contextual dimension relates intelligence to the external work of the individual. It emphasizes adaptation to, selection of, and shaping of the environment. Sternberg has compared his information-processing theory of intelligence with the hierarchical theories of intelligence and believes that a general intelligence, *g*, accounts for the correlations among scores on different tests. He explains g primarily – though not entirely – in terms of metacomponents.

Sternberg offered another theory of intelligence, referred to as successful intelligence, that focuses on "the ability to adapt to, shape, and select environments to accomplish one's goals and those of one's society and culture" (Sternberg & Kaufman, 1998, p. 494). His successful intelligence theory complements his triarchic theory and suggests that individuals with successful intelligence are able to discern their strengths and weaknesses and then determine ways to use their strengths and minimize their weaknesses. The three broad areas associated with successful intelligence are analytical, creative, and practical (Sattler, 2001). Underlying the theory is the premise that schools do not use children's multiple abilities. The theory emphasizes the importance of other aspects of intelligence and their usefulness in our society. However, these aspects are not typically measured well by standardized intelligence tests.

Jean Piaget perceived intelligence as a form of biological adaptation to one's environment. According to Piaget, two inherent tendencies govern interactions with the environment: organization and adaptation. Adaptation contains two complementary processes: assimilation and accommodation. Piaget's model of intelligence is a hierarchical one, in which cognitive development is divided into four major periods, each characterized by stages and substages. While there is presently no comprehensive battery of Piagetian test of intelligence, studies have found positive correlations between Piagetian measures and psychometric scales of intelligence in infant, preschool, and school-age populations (Sattler, 2001).

The factor analytic theories of intelligence presented above represent a significant departure from traditional views and conceptualizations of the structure of intelligence. Although the theories have undergone varying degrees of empirical validation, they present viable foundations from which to construct new measures of intelligence – measures that may lead to greater insights into the nature, structure, and neurobiological structures of cognitive functioning and that may be more appropriate for assessing the cognitive abilities of individuals from culturally, linguistically, and ethnically diverse backgrounds.

Cattell and Horn's Fluid and Crystallized Theory of Intelligence

Over the middle decades of the 20th century, intelligence tests themselves changed relatively little. This period could be characterized as a period of consolidation of the methods that had been developed in the first 25 years (Thorndike, 1997). The tests produced by the factor analysts explored some new content areas, but mostly without fruitful results. The Wechsler scales and the revisions of the Stanford-Binet broke no new ground. A single bright spot in the area of theory is found in the world of Raymond B. Cattell (1963) and his student, John Horn (Horn & Cattell, 1967; Horn, 1985) with the theory of fluid and crystallized intelligence (*Gf* – G*c* theory of intelligence).

G*f* - G*c* theory was first postulated by Cattell (1941, 1957) to consist of two major types of cognitive abilities. Cattell considered fluid intelligence (G*f*) to include inductive and deductive reasoning, abilities that were thought to be influenced primarily by biological and neurological factors, as well as by incidental learning through interaction with the environment (Taylor, 1994). In contrast, crystallized intelligence (G*c*) was postulated to consist primarily of abilities, especially knowledge, that reflected individual differences due to the influences of acculturation (Gustafsson, 1994; Taylor, 1994). Thus, the original G*f* - G*c* theory was a dichotomous conceptualization of human cognitive ability. The G*f* - G*c* theory label has been retained as the acronym for this theory despite the fact that the theory has not been conceived of as a dichotomy since the 1960s (Gustafsson & Undheim, 1996; Horn & Noll, 1997; Woodcock, 1990). As a result, G*f* - G*c* theory is often misunderstood as being a two-factor model of the structure of intelligence.

As early as the mid-1960s, Horn (1968) expanded the G*f* - G*c* model to include four additional cognitive abilities in the domains of visual perception or processing (G*v*), short-term memory (G*sm*), long-term storage and retrieval (G*lr*), and speed of processing (G*s*). Horn (1968) next refined the definitions of G*v*, G*s*, and G*lr*, and added an auditory processing ability (G*a*). More recently, factors representing a person's quantitative ability or knowledge (G*q*) and facility with reading and writing (G*rw*) (Horn, 1985, 1988a, 1991; Woodcock, 1994) were added to the model, resulting in a 9-factor ability structure. Historically, this body of research has been delineated G*f* - G*c* theory.

Following an extensive review of reanalysis of most of the theoretical and empirical research on human cognitive abilities and their measurement since the late 1900s, Carroll (1993, 1997) proposed a “Three-Stratum Theory of Cognitive Abilities” that, according to a number of scholars, is a benchmark and the most ambitious attempt to provide a psychometric taxonomy of intelligence abilities to date (Gustafsson & Undheim, 1996).

Carroll’s secondary factor analyses provided evidence for a hierarchical structure of human intelligence, which included a general intelligence factor (*g*), broad abilities, and narrower abilities subsumed under these broad abilities. His model organizes cognitive ability at three strata that differ as a function of breadth or generalizability of abilities. Within this conceptualization, the 69 specific narrow abilities consume the base of the model, known as stratum I. Narrow abilities “represent greater specializations of abilities, often in quite specific ways that reflect the effects of experience and learning, or the adoption of particular strategies of performance” (Carroll, 1993, p. 634). The broad abilities are located in the middle layer, known as stratum II. According to Carroll (1993), broad abilities represent “basic constitutional and long standing characteristics of individuals that can govern or influence a great variety of behaviors in a given domain” and they vary in their emphasis on process, content, and manner of response (p. 634). The general intelligence factor or *g* represents the pinnacle of the model, known as stratum III. Each of the broad and narrow abilities within the structure loads on this general intelligence factor or *g*.

Psychometric intelligence theories have converged recently on the more "complete" G*f* - G*c* multiple intelligences taxonomy, reflecting a review of the extant factor-analytic research conducted over the past 60 years (McGrew & Flanagan, 1998). The Cattell-Horn-Carroll (CHC) theory constitutes an integration of these independent empirical sources (i.e., Cattell-Horn G*f* - G*c* theory and the Carroll three stratum theory). Within the CHC conceptual framework, overall cognitive functioning is differentiated hierarchically among the distinct broad abilities and narrow abilities consolidated from Cattell/Horn and Carroll's work, respectively. At the broadest level (Stratum III) is a general intelligence factor conceptually similar to Spearman's *g*. Next in breadth are nine broad abilities (Stratum II), including processing speed (G*s*), short-term acquisition and retrieval (G*sm*), fluid intelligence (G*f*), long-term storage and retrieval (G*lr*), broad auditory perception (G*a*), broad visual perception (G*v*), crystallized intelligence [acquired knowledge] (G*c*), quantitative knowledge (G*q*), and reading/writing ability (G*rw*). Subsumed under these broad Stratum II abilities are 69 first order narrow abilities (Stratum I abilities). For example, phonetic coding (PC), a narrow ability of broad auditory perception (Ga), is the capacity to process speech sounds or phonemes, as in identifying, isolating, and blending sounds.

In the theory, a continuum is posited where certain broad abilities are designated as less susceptible to formal learning/environmental exposure and other broad abilities are designated as more permeable to environmental exposure. McGrew and Flanagan have described this continuum in the following way, "each of the broad cognitive abilities can be thought of as lying on a continuum, with abilities that depend little on direct instruction and formal learning (e.g., G*f*) at one end and abilities that depend extensively on breadth and depth of knowledge or culture, including the ability to communicate (especially verbally) and reason through the application of previously learned procedures (e.g., G*c*), at the other. The remaining CHC broad abilities lie somewhere along this continuum, with their location depending on the degree to which they differ as a function of relative emphasis on process, content, and manner of response" (p. 85).

At the more inherent end of the continuum are low-level processing abilities and high-level processing abilities. Low-level processing abilities constitute the capacities necessary for automatic and efficient processing of straightforward information. The broad abilities included at this level are short-term memory (G*sm*) and processing speed (G*s*). High-level processing abilities, often referred to as "thinking abilities," constitute those abilities employed during novel learning and problem solving tasks. During such tasks a variety of cognitive processes can be elicited including reasoning (G*f*), storage of new information and retrieval of previously learned information (G*lr*), perception and processing of auditory stimuli (G*a*), and perception and processing of visual stimuli (G*v*). At the opposite end of the continuum (less inherent end) are acquired knowledge abilities or those abilities largely dependent upon formal instruction, prior experience, and exposure. Abilities in this area include crystallized ability (G*c*), quantitative reasoning (G*q*), and reading/writing ability (G*rw*).

Grounded in more than half a century of confirmatory and exploratory factor-analytic research, the amalgamated CHC theory of cognitive abilities is considered to be one of the most comprehensive and empirically validated theories of cognitive functioning to date (Evans, Floyd, McGrew, & Leforgee, 2002; McGrew & Flanagan, 1998).

Lineal descendants of the early intelligence tests, now greatly modified and more culturally fair than the early tests, are widely used in schools, institutions, the military, and industry. Despite variations in the definitions of intelligence and differences among theorists about how intelligence is organized, the advancement and perseverance of mental measurement over the past century has proven that it is useful, is beneficial to society, and remains one of psychology's major contributions to modern life in America.

The Predictive Value of Intelligence Testing

IQ prediction of achievement. It is commonly believed that one's level of overall cognitive ability (e.g.., global IQ; *g*) represents his or her potential for academic achievements (or academic success) (Flanagan, Ortiz, Alfonso, & Mascolo, 2002). Jensen (1985), Herrnstein and Murray (1994), and Gottfredson (2002a, 2002b) asserted that cognitive ability, as measured by a full scale IQ score, is the basis for a person's educational level, job attainment and performance, and socioeconomic status. Herrnstein and Murray claim that after holding environmental effects constant, a person with low IQ is most likely to drop out of high school, not gain a higher education, have children at a younger age, have marriages that fail, accept welfare, be unemployed, and live below the poverty level. Flanagan, Andrews, and Genshaft (1997) added that IQ tests have predictive ability in that they can estimate a person's current and future academic performance and that they have treatment validity in that they can be used to help create an appropriate educational program for the child.

Assessment of academic abilities or achievement in areas such as reading and math is typically accomplished through the use of standardized, norm-referenced tests of achievement. Yet, depending on the specific instruments chosen for assessment, certain academic abilities appear to be measured by intelligence tests, whereas certain cognitive abilities appear to be measured by achievement tests.

According to Horn (1988b), “Cognitive abilities are measures of achievements, and measures of achievements are just as surely measure of cognitive abilities” (p. 655). Carroll (1993) echoed this conceptualization when he stated, “It is hard to draw the line between cognitive abilities and cognitive achievements. Some will argue that all cognitive abilities are in reality learned achievements of one kind or another. Such an argument is difficult to counter, because it is obvious that the performances required on even the most general tests of intelligence depend on at least some learning” (p. 510).

Rather than conceiving of cognitive abilities and academic achievements as independent, they may be better thought of as lying on an ability continuum that has the most general types of abilities at one end and the most specialized types of knowledge at the other, the latter of which develops more through an individual’s instructional and education experiences (Carroll, 1993). Through an examination of the two broad cognitive abilities (G*f* and G*c*), G*f* represents processing or thinking abilities that develop largely independent of formal education and school related experiences, whereas, G*c* abilities reflect more specialized types of knowledge that develops largely as a function of formal education and direct learning and instruction. Therefore, rather than being mutually exclusive, academic and cognitive abilities may be best conceived as of as lying of the continuum as discussed by Carroll.

The Relationship between Cognitive Abilities and Achievement

Commonly used individually administered intelligent test batteries provide reliable and valid estimates (e.g., Full Scale IQ) of an underlying general ability construct (psychometric *g*) that is typically interpreted as an individual's global level of cognitive functioning. Considerable empirical evidence indicates that this general ability is "among the most dominant and enduring factors, both causal and corollary, associated with scholastic occupational success; environment adaptation; physical propensity and morbidity; and scientific, cultural, and political acumen" (McDermott, Fantuzzo, & Glutting, 1990, p. 291).

A review of the cognitive abilities research literature reveals that attempts to move "beyond *g*" (i.e., the addition of specific abilities to g in the prediction and explanation of educational and occupational outcomes) have been met with mixed results. In his presidential address to the American Psychological Association (APA), McNemar (1964, p. 875) concluded, "The worth of the multitest batteries as differential predictors of achievement in school has not been demonstrated." Cronbach and Snow (1977) reached a similar conclusion in their review of the aptitude-treatment interaction (ATI) research, which demonstrated that interventions interact primarily with general levels of intelligence, and that few, if any, meaningful specific ability-treatment interactions exist. Jensen (1984, p. 101) also reinforces this conclusion when he stated that "*g* accounts for all of the significantly predicted variance; other testable ability factors, independent of *g*, add practically nothing to the predictive validity." In the area of intellectual assessment, the failure to establish the importance of specific abilities has resulted in the argument against the practice of interpreting subtest scores in individual batteries (McDermott, Fantuzzo, & Glutting, 1990; McDermott & Glutting, 1997; Glutting, Watkins, Konold, & McDermott, 2006). The inability to move beyond *g* has provided little optimism for the development of interventions designed according to an individual's profile of specific ability strengths and weaknesses (Vanderwood, McGrew, Flanagan, & Keith, 2002). However, attempts to move beyond g have recently begun to bear fruit (McGrew, Flanagan, Keith, & Vanderwood, 1997).

Despite the failure to demonstrate the importance of specific cognitive abilities, Carroll (1993, p. 676) concluded, “There is no reason to cease efforts to search for special abilities that may be relevant for predicting learning.” McGrew, Flanagan, Keith and Vanderwood (1997), Flanagan (2000), and Keith (1999) provided support for this position when they suggested that recent advances in theories of intelligence, applied theory-driven measurement of intelligence, and research methodology (e.g., structural equation modeling or SEM) argue for continued efforts to investigate the effects of general and specific abilities on general and specific achievements.

A variety of intelligence theories grounded in markedly different research traditions have received increased attention in recent years (e.g., CHC theory of cognitive abilities, Gardner’s Theory of Multiple Intelligences, the Luria-Das Model of Information Processing, Sternberg’s Triarchic Theory of Intelligence). Of these theories, the psychometrically based CHC theory has been viewed as having the greatest potential for examining the importance of general and specific cognitive abilities, as it is based on a more thorough network of validity evidence than other contemporary multidimensional ability models of intelligence (McGrew & Flanagan, 1998).

Cognitive Abilities and Reading Achievement

Crystallized intelligence (Gc). Crystallized ability refers to information learned through previous experience. Crystallized ability, an area of acquired knowledge, includes one’s cultural knowledge obtained through exposure and one’s ability to communicate this knowledge. Crystallized ability also includes the capacity to reason with previously learned procedures. This reasoning depends on previously learned information either encountered through formal schooling or in interactions with the environment (Flanagan, Ortiz, Alfonso, & Mascolo, 2002; McGrew & Flanagan, 1998).

Studies using CHC theory suggest that psychometric *g* (i.e. the general factor that results from the positive correlation of mental tasks) and five specific cognitive abilities significantly contribute to variance on measures of reading achievement. Important specific cognitive abilities include G*a* (auditory processing), G*c* (crystallized intelligence or knowledge), G*lr* (long term storage and retrieval), G*sm* (short-term memory) and G*s* (processing speed) (Benson, 2007; Evans, Floyd, McGrew, & Leforgee, 2002; Flanagan, 2000; Floyd, Keith, Taub, & McGrew, 2007; Hale, Fiorello, Kavanagh, Hoeppner, & Gaither, 2001; McGrew, Flanagan, Keith & Vanderwood, 1997; Vanderwood, McGrew, Flanagan, & Keith, 2002). The present study focuses on fluid reasoning (G*f*) and crystallized ability (G*c*) and their relationship to academic achievement. Therefore, these areas will be the focus when discussing prior research findings correlating reading achievement and cognitive processes.

McGrew, Flanagan, Keith, and Vanderwood (1997) examined the relationship between *g* and seven G*f* - G*c* specific abilities and general and specific reading skills. This study was designed to reexamine the g versus specific abilities issues in a manner that reflected progress in theory, measurement and methodology. Analyses were conducted in separate model calibration and cross-validation samples at each of five grade levels. This study used structural equation modeling procedures and tests from a validated G*f* - G*c* organized intelligence battery to operationalize a hierarchical *g* and G*f* - G*c* model consistent with Carroll's (1993) three stratum model of intelligence.

Across all analyses, the relationship of *g* to general reading was as expected – significant and strong across all developmental levels, ranging from .57 to .88. However, a number of significant and strong cross-validated direct effects for specific G*f* - G*c* abilities on specific reading skills also were found, as a result of the utilization of SEM analyses. McGrew et al. (1997) found that Ga abilities were significantly related to Word Attack skills (pronunciation or decoding of unfamiliar words). The *g* factor displayed a significant indirect effect on Word Attack skills (as mediated through its strong direct effect on reading), which in turn had a consistently strong direct effect on G*a* abilities contributed to the explanation of Word Attack skills as reflected by the significant direct effects ranging from approximately .20 to .50 in grades 1 through 9. Thus both general (*g*) and specific abilities (G*a*) were found to be important for understanding Word Attack skills.

The results for Passage Comprehension (a person's skill in identifying a key word missing from a reading passage) reflect the importance of a different specific ability (G*c*) and reveal potentially important developmental trends. Although the effects of *g*, general reading, and G*c* were all significant during grades 1 – 6, the general effects (i.e., *g* and general reading) decreased in importance and became nonsignificant after grade 6, whereas the specific G*c* ability effect gradually increased with age. Interestingly, G*c* was the strongest of all effects on Passage Comprehension at all grades.

Overall, the results of the McGrew et al. (1997) study suggest that some specific abilities, including crystallized intelligence and auditory processing, may be important for understanding the development of specific reading skills, above and beyond the understanding gained from general cognitive (*g*) and achievement constructs. In addition, the relative importance of both general and specific abilities change developmentally, and these changes may in turn vary as a function of the specific academic skill being investigated.

Flanagan (2000) investigated the relationship between Wechsler-based CHC Cross-Battery Assessment and reading achievement. Her findings were two-fold: (1) the *g* factor underlying the Wechsler-based CHC cross-battery model accounted for a substantial variance in reading achievement, up to 68% and (2) when assessments are organized around the strong CHC theoretical model, specific cognitive abilities, including auditory processing (G*a*), crystallized intelligence (G*c*), and processing speed (G*s*) explained a significant portion of variance in reading achievement beyond that accounted for by *g*.

Evans, Floyd, McGrew, and Leforgee (2002) examined the relations between the CHC theory of cognitive abilities and reading achievement during childhood and adolescence. Using a large, nationally representative sample including students 6 to 19 years of age, operational measures of CHC cognitive abilities obtained from the Woodcock Johnson III (WJ III) were found to be significantly related to the components of reading achievement. Multiple regression analyses were used to regress several WJ III cognitive clusters onto the WJ III Basic Reading Skills and Reading Comprehension clusters for 14 age groups. General knowledge (G*c*) or crystallized ability demonstrated moderate to strong relations with components of reading achievement across childhood and adolescence, and short-term memory (G*sm*) demonstrated moderate relations throughout this period. Auditory processing (G*a*), long-term retrieval (G*lr*), and processing speed (G*s*) demonstrated moderate relations with the components of reading achievement during the elementary school years. In contrast, fluid reasoning (G*f*) and visual-spatial thinking (G*v*) demonstrated no consistent pattern of significant relations across childhood and adolescence. Although previous analyses examining the predictive power of the Woodcock-Johnson Psychoeducational Battery – Revised (WJ-R; Woodcock & Johnson, 1989) – an empirically supported measure of several constructs within the Cattell-Horn-Carroll (CHC) theory of cognitive abilities – G*f* cluster indicated that G*f* abilities are significantly related with Reading Comprehension from the early school-age years to early adulthood (McGrew, 1993), the results of the Evans et al. study suggest that, within the context of all seven CHC factor clusters, G*f* does not appear to add anything unique or significant to the prediction of reading achievement.

Using the G*f* - G*c* model factor scores in predicting concurrent standardized academic achievement test performance in a clinic-referred sample of children with diagnosed learning disabilities, Hale, Fiorello, Kavanagh, Hoeppner, and Gaither’s (2001) results revealed complex relationships among collinear predictors, with crystallized, quantitative and short-term memory factors accounting for most achievement variance, regardless of academic domains. The G*c* factor appeared to predict consistently all academic domains, which is not surprising, again considering that achievement measures are related to crystallized abilities (McGrew & Flanagan, 1998) or acquired knowledge (Kamphaus, 1993), abilities inseparable from previous experience and education. Of importance to note, commonality analyses were employed, as this method has become the preferred analyses over hierarchical regression models

(Pedhazur, 1997) when attempting to predict academic achievement, particularly when factor or subset scatter is large, as it was in Hale et al.'s sample.

In a study conducted by Vanderwood, McGrew, Flanagan, and Keith (2002), the standardization sample of the WJ-R was used to analyze the contribution of specific cognitive abilities to reading achievement at five developmental levels. The results of the study clearly indicate that G*c* (comprehension knowledge or crystallized ability) and G*a* (auditory processing) play an important role in the development of reading skills, particularly reading comprehension and basic reading skills, respectively. The results of this study support the belief the CHC specific cognitive abilities can be used to explain and better understand academic achievement, above and beyond the effects of *g*.

Floyd, Keith, Taub, and McGrew (2007) employed structural equation modeling to examine the effects of CHC abilities on reading decoding skills using five age-differentiated samples from the standardization sample of the WJ III. Using the Spearman Model including only *g*, strong effects of *g* on reading decoding skills were demonstrated at all ages ranging from .73 - .88. Using the Two-Stratum Model including *g* and broad abilities, direct effects on the broad abilities of G*lr*, G*s*, G*c*, G*sm* and G*a* on reading decoding skills were demonstrated at select ages. *G* had very large but indirect effects (ranging from .64 - .81), meaning that *g* had direct effects on the broad ability factors, and in turn, some of these broad ability factors had direct effects on reading decoding skills. Important to note, beginning at ages 7 to 8 and continuing through the three remaining age levels, G*c* demonstrated large direct effects.

Benson (2007) applied structural equation modeling procedures to the standardization sample of the WJ III to simultaneous estimate the effects of a psychometric general factor (*g*), specific cognitive abilities, and reading skills on reading achievement. The results of this study indicated that g had a strong direct relationship with basic reading skills until about sixth grade, at which point *g* ceases to have a direct effect but continues to have an indirect effect mediated through G*c* and G*sm*. Also, *g* was found to have a strong indirect effect on reading fluency and comprehension across grade levels. The influence of G*c* on basic reading skills was minimal at Grades kindergarten to 3 and 4 to 6 and moderate to strong at Grades 7 to 12. Benson noted that this may be related to the improvement of reading strategies as children age. That is, older children may develop strategies allowing them to better utilize their background knowledge. The results of this study suggest that the influence of G*c* on reading comprehension increases as children develop basic reading skills. This finding is consistent with the results of previous research (Yavanoff, Duesbery, Alonzo, & Tindal, 2005) and may reflect a bidirectional relationship between G*c* and reading comprehension. Thus, acquired verbal knowledge may improve reading comprehension, which in turn may stimulate the development of G*c*.

In relation to the present study, prior research suggests that *g* and specific cognitive abilities play important roles in reading achievement. *G* is shown to have direct (Benson, 2007; Flanagan, 2000; Floyd et al., 2007; McGrew et al. 1997) and indirect effects (Benson, 2007; Floyd et al., 2007; McGrew et al. 1997) on all aspects of reading ability, including word attack skills, reading decoding skills, basic reading skills, reading fluency and reading comprehension. G*c* has been found to have a moderate to strong relationship with reading achievement in general (Evans et al., 2002; Flanagan, 2000; Hale et al., 2001) and with more specific reading skills including reading decoding skills

(Floyd et al., 2007) and reading comprehension (McGrew et al. 1997; and Vanderwood et al., 2002).

It is not surprising that G*c* is generally a strong predictor of Basic Reading Skills and Reading Comprehension. The relations between reading achievement and the breadth and depth of a person's knowledge are logical. Both abilities stem largely from the acquisition of declarative and procedural knowledge and, in fact, may be considered types of academic achievement. It is clear that the link between G*c* and reading achievement is robust and increasing as a function of age. This link may reflect a bidirectional relationship, whereas vocabulary and general knowledge contribute to reading abilities and vice versa (Stanovich, 1986).

The significant relationship between G*c* and reading comprehension is consistent with research that indicates prior achievement affects future achievement and that comprehension-knowledge influences reading achievement. It also seems logical that prior exposure to the culture and language of the test should be of benefit to the test taker when attempting to garner understanding from a reading passage or define vocabulary. (Vanderwood et al. 2002).

Cognitive Abilities and Math Achievement

Fluid intelligence (Gf). Fluid intelligence refers to mental operations that an individual uses when faced with relatively novel tasks that cannot be performed automatically. These mental operations may include forming and recognizing concepts, perceiving relationships among patterns, drawing inferences, comprehending implications, problem solving, extrapolating, and reorganizing or transforming information. Inductive and deductive reasoning are generally considered to be the hallmark narrow-ability indicators of G*f*. Although most practitioners would agree that this ability is typically not measured directly by individually administered achievement batteries, some tests of achievement clearly involve the use of G*f* abilities. For example, tests of reading comprehension may require individuals to draw inferences from the text (Flanagan, Ortiz, Alfonso, & Mascolo, 2002).

Aside from general inductive and deductive reasoning abilities, G*f* also subsumes more specific types of reasoning, most notably quantitative reasoning (RQ). Quantitative reasoning is directly related to formal instruction and classroom related experiences. In many ways, it can be seen as a measure of learning, and in fact, subtests that measure mathematics reasoning on a number of achievement tests (Wechsler Individual Achievement Test – Third Edition [WIAT-II] and WJ-3) appear to directly measure RQ. As such, RQ seems to lie well within the overlap of the continuum of cognitive and academic abilities. At the very least, it represents a good example of one of those abilities that is often measured by both cognitive and achievement tests (Flanagan et al., 2002). In comparison to reading-related competencies, relatively little is known about the development and maintenance of mathematics skills. Less is known about the underlying cognitive processes that contribute to mathematics achievement and mathematical disabilities (Rourke & Conway, 1997).

Several studies have examined the relationship between measures of CHC broad and narrow cognitive abilities and mathematics achievement. For example, using multiple regression analyses of the WJ-R G*f* - G*c* clusters and a group-administered achievement test with a sample of school-age children, Williams, McCallum and Reed (1996) reported that G*f* and G*c* were the best predictors of mathematics achievement.

McGrew and Hessler (1995) examined the relations between the seven WJ-R G*f* - G*c* cognitive clusters and the WJ-R Basic Mathematics Skills and Mathematics Reasoning clusters. Fluid reasoning and crystallized ability correlated consistently and significantly with mathematics achievement, above and beyond the contribution of *g*. The G*c* relation with mathematics achievement increased monotonically with age. Fluid reasoning (G*f*) generally displayed moderate relations with both mathematics clusters across all ages.

McGrew et al. (1997) examined the relationship between *g* and seven G*f* - G*c* specific abilities and general and specific mathematics skills. This study was designed to reexamine the *g* versus specific abilities issues in a manner that reflected progress in theory, measurement and methodology. Analyses were conducted in separate model calibration and cross-validation samples at each of five grades levels. This study used structural equation modeling procedures and tests from a validated G*f* - G*c* organized intelligence battery to operationalize a hierarchical *g* and G*f* - G*c* model consistent with Carroll’s (1993) three stratum model of intelligence.

Across all analyses, the relationship of *g* to general math was as expected – significant and strong across all developmental levels, consistently ranging from .43 to .56, with the exception of a lower .23 effect at grades 5-6. The indirect and direct effects for the general constructs (*g* and general mathematics) increased in importance with age. In contrast, the specific abilities of G*f* and G*c* decreased in relative importance with increasing age. An interesting developmental observation was the finding that G*f* abilities were much more strongly associated with solving applied mathematics problems at grades 1-2 than any other general or specific ability, with its importance decreasing and becoming nonsignificant after grade 9.

Similar findings have surfaced from independent research using the WJ-R G*f* - G*c* clusters and from additional analyses of the WJ-R standardization sample that included *g* in the analysis. Keith (1999) found that measures of mathematics achievement were strongly influenced by *g* and a number of other CHC broad cognitive abilities, such as G*c*, G*f*, and G*s*. Hale et al. (2001) reported that when CHC-organized commonality analyses were applied to factor indexes, the broad cognitive ability factors G*c*, G*sm*, and G*q* were significantly associated with mathematics achievement above and beyond the predictive effects of the Full Scale IQ.

Floyd, Evans, and McGrew (2003) used multiple regression analyses to investigate the validity of the cognitive clusters from the WJ III in predicting mathematics calculation and mathematics reasoning skills. Comprehension-knowledge (G*c*) demonstrated the strongest predictor of mathematics achievement throughout the school-age years, indicating moderate relations with Math Calculation Skills after the early school-age years and moderate to strong relations with Math Reasoning. General cultural knowledge and knowledge of mathematics concepts, facts, and the procedures to conduct arithmetic stem largely from the acquisition and modification of declarative and procedural knowledge structures (Woodcock, 1993, 1998). Both abilities may be considered types of academic achievement (Anastasi, 1988; Flanagan et al., 2002; Kaufman, 1994; Woodcock, 1990). Because of these similarities, the increasing strength of relations between G*c* and Math Reasoning are logical because knowledge of mathematics, rather than more fundamental cognitive processes, likely contributes significantly to further mathematics skill development after basic mathematics skills (e.g., simple addition and multiplication) are established (Geary, 1994). Furthermore, because G*c* subsumes narrow cognitive abilities associated with listening and speaking, the moderate to strong relations between G*c* and the mathematics clusters are likely due to the influence of language-based cognitive processes on math performance.

In the same study, Floyd et al. (2003) found that Fluid Reasoning (G*f*) generally demonstrated moderate relations with the Math Calculation Skills and moderate to strong relations with Math Reasoning throughout childhood and adolescence, indicating that domain-general problem-solving and reasoning abilities are strong influences on mathematics achievement. The authors list several reasons as to why this finding is not surprising. First, Carroll's (1993) analysis and comprehensive review of human cognitive abilities indicated that quantitative reasoning abilities could be included under the stratum II ability, Fluid Intelligence. Second, G*f* appears to represent some of the prominent constructs in studies of mathematics skill development, such as problem-solving schemata, strategy use, and strategic change (Cummins, 1991; Lemaire & Siegler, 1995; Vaughn & Wilson, 1994). Third, research that has included instruments that appear to measure G*f* abilities (e.g., the Wisconsin Card Sorting Test and the Halstead Category Test) has indicated that such abilities are significant correlates of mathematics achievement (Bull, Johnston, & Roy, 1999; Rourke, 1993; Shute & Huertas, 1990; Strang & Rourke, 1983).

Proctor and Shaver (2005) extended previous research examining the relationship between CHC cognitive abilities and math achievement. The cognitive profiles of children with normative weaknesses in Math Calculation Skill or Math Reasoning were compared to those of their average-achieving peers. The cognitive profile of the low Math Calculation Skills group was similar to that of their average-achieving peers. The low Math Reasoning group scored lower than their average-achieving peers on the cognitive abilities as a set, which may suggest group differences in *g*. As a group, their scores on fluid reasoning and comprehension-knowledge were lower than those of the Average Achievement group. In addition, their mean score in the fluid reasoning cluster was significantly lower than the population mean, indicating a normative weakness. This relationship between math reasoning and fluid reasoning resonates with the findings of previous studies examining the statistical relations between math skills and underlying cognitive abilities of reasoning and novel problem-solving skills.

The studies reviewed above provide evidence of the external validity of CHC broad and narrow cognitive abilities in predicting mathematics achievement, even when the effect of *g* is present in the analysis. *G* is shown to have direct effects (Keith, 1999; McGrew et al., 1997) on math achievement in general. G*c* has been found to have a moderate to strong relationship with math achievement in general (Hale et al., 2001; Keith, 1999; Williams et al., 1996) and with more specific math skills including Math Calculation (Floyd et al., 2003) and Math Reasoning (Floyd et al., 2003; McGrew et al., 1997; Proctor & Shaver, 2005). G*f* has been found to have a moderate to strong relationship with math achievement in general (Keith, 1999; Williams et al., 1996) and with more specific math skills including basic skills (McGrew et al., 1995), math calculation (Floyd, et al., 2003) and math reasoning (Floyd et al., 2003; McGrew et al., 1995 and 1997; Proctor & Shaver, 2005).

In sum, recent research has demonstrated that specific cognitive abilities are important in understanding specific academic skills. Furthermore, these specific abilities explain a significant portion of the variance in academic criteria (reading and math achievement) above and beyond that accounted for by *g*, or general ability. The importance of these specific abilities in understanding reading and mathematics achievement, in particular, has been found to be indistinguishable across ethnic groups (Keith, 1999). In general, this research shows that specific cognitive abilities (G*f*, G*a*, G*s*, G*c*, G*lr*, and G*sm*) have significant, cross-validated effects on the development of specific reading and math skills (McGrew et al., 1997). Also, these cognitive abilities are significantly related to academic skills at different developmental levels. This research suggests that certain cognitive abilities, at certain ages, may be important to measure and interpret in addition to *g* (or general intelligence) in reading- and math-related referrals. The results of these studies further suggest that the momentum of the anti-specific ability and anti-intelligence tests pendulum should be slowed or even reversed.

Cognitive Abilities and Prediction of Scholastic Achievement

Considering the movement toward using the CHC theory of cognitive abilities to understand students' cognitive strengths and weaknesses, it follows that one would examine the predictive value of cognitive scores on achievement. Scholastic achievement can be measured two ways: performance on standardized achievement tests and actual grades received. While research has compared cognitive ability scores with an overall grade point average (GPA), with the correlation between IQ scores and grades at about .50 (Neisler et al., 1996; Sternberg, Grigorenko, & Bundy, 2001), until this point, most research has focused on the comparison between cognitive ability scores with scores on standardized achievement tests (Evans, et al., 2002; Felton & Pepper, 1995; McBride-Chang, 1995; McGrew & Flanagan, 1998; Wagner & Torgesen, 1987; Williams, McCallum, & Reed, 1996). One specific example taken from previous research shows a relationship between crystallized ability (G*c*) and reading achievement as measured on standardized achievement tests (Felton & Pepper, 1995; McBride-Chang, 1995; Wagner & Torgesen, 1987). However, there is little research that indicates that above average ability in G*c* correlates with higher grades in reading, language arts or English classes. McLaughlin (2005) studied the relationships between specific cognitive factors within the CHC Theory and grades as markers of achievement in specific subjects in a sample of high school and college students. Results did not demonstrate significant correlations among CHC factors of cognitive abilities and high school grades or college GPA. As McLaughlin notes, because of the small sample size (N = 30), results should be interpreted with caution. The current study expanded this investigation to look at the correlations between cognitive ability and grades in school, as grades are the typical measure of scholastic achievement as defined by the educational system. Rather than focusing on an aggregate point system, such as GPA, grades attained in specific subject areas (i.e., reading, writing, math, science and social studies) were examined individually.

Standardized Achievement Tests and Prediction of Scholastic Achievement

Standardized, selected-response achievement tests are the most common method for measuring student performance. Each year, millions of individuals in the United States take standardized achievement and post-secondary admissions tests (e.g., SAT [formerly the Scholastic Aptitude Test and Scholastic Assessment Test]). Given their prominent role in influencing educational opportunities, these tests are of great interest to the public. Research indicates that scores on standardized tests of ability, such as the SAT, as well as past academic performance (generally measured by high school grade point average and class rank) are the most valid predictors of success in college, as measured by college GPA. When corrections for measurement unreliability and range restriction are taken into account, scores on standardized tests have demonstrated strong criterion-related validities with cumulative college GPA (r = .45) and correlations with high school grade point average and rank (rs = .44 to .62); correlations with first-year college GPA are often higher (Schmitt et al., 2007). While it is concluded that achievement test results are generally good predictors of grade point averages in high school and college samples (Kuncel, Hezlett & Ones, 2001, 2004), few research studies have looked at the relationship between achievement test results and actual grades earned in specific subject areas, rather than GPA. For example, it is unclear as to whether strong performance on a verbal section of an achievement test translates into higher grades in the subject areas of reading and writing. The current study expanded this investigation by looking at the correlations between achievement test results and grades in a middle school sample, as grades are the typical measure of scholastic achievement as defined by the educational system. Rather than focusing on an aggregate point system, such as GPA, grades attained in specific subject areas (i.e., reading, writing, math, science and social studies) were examined individually.

Group-Administered Tests

In the late 19th century, Alfred Binet's developments in the field of intelligence testing did not occur in a vacuum. American psychologists also were in active pursuit of a measure of mental ability. The breakthrough in mental measurement came about as a result of the entry of the United States into the First World War. All versions of the Binet scales had to be given by a trained technician to one person at a time. But group testing, where subjects read questions to themselves and check off multiple-choice answers or make appropriate marks on the form, would be far quicker, simpler, and very much less expensive.

Psychological testing got a strong boost from the American Psychological Association (APA) in 1895, with the appointment of a committee on testing (Thorndike, 1997). The committee reported that the most useful contribution of the profession would be the development of psychological examinations that could be quickly given to large numbers of military personnel so as to eliminate the mentally incompetent, classify individuals according to their abilities, and select the most competent for special training and responsible positions. In 1899, Kirkpatrick (1900), addressing the APA, called for tests "of such nature that they can be taken by children as well as adults, that they shall be such that all persons tested will have had about an equal opportunity for the exercise of the power tested, and that in the interest of economy of time the tests so far as possible shall be so planned that they can be given to a whole class or school at once" (pp. 279-280).

Lewis Terman, Henry Goddard, and Robert Yerkes began planning the tests. Yerkes was commissioned a major in the Army and produced the Army Alpha, a written test of intelligence, and the Army Beta, a pictorial test for the functionally illiterate. By the time the war ended in 1918, more than 1.7 million men had taken the tests and over three hundred psychologists had graded each man and suggested a suitable military assignment (Hunt, 1993).

After the war, the testing movement took off. As early as 1909, Binet called for universal measurement of the intelligence of school pupils. In 1920, as the first group-administered tests for use in the schools were being published, he stated “the greatest usefulness [of intelligence tests] will be found in their universal application to school children……‘A mental test for every child’ is no longer an unreasonable slogan” (Terman, 1920, p. 20). As a start, he proposed that all students in the fourth grade and beyond should be give a group-administered test every year and those whose scores were at the extremes (very high or very low) should be given an individually administered Binet test. As a result, the use of intelligence tests in schools grew rapidly, with approximately a million children in the schools of the United States given a group mental test during the year 1919-1920 (Thorndike, 1997).

In deciding whether to use group or individually administered tests, Sattler (2001) recommends that one must consider the nature of the referral question and whether the referral question can be answered without administering a battery of individual tests. When there is a reason to question the validity of the results of group tests or when an observation of the examinee’s performance is needed, the administration of individual tests is recommended.

Group tests are less frequently used in the assessment of special children for four reasons (Sattler, 2001). First, group tests usually require some degree of reading proficiency, and many children with special needs have reading difficulties. Second, because examinees taking group tests typically fill in bubbles, circle letters or underline answers instead of giving their answers orally, it is difficult to determine whether they know the answers or are merely guessing. Further, group-administered tests do not allow for the observation of the examinee's problem solving strategies. Third, group tests tend to use recognition rather than recall, requiring examinees to select one answer from among the several that are given. Fourth, examinees taking group tests can get lost, bored, fatigued, or indifferent without examiners' knowing that such behaviors are present or having a chance to intervene. With individual tests, examiners can monitor these factors and take steps to reduce their influence by providing encouragement, bringing examinees back to focus, and taking short breaks.

Individually administered tests are more expensive in both cost and time than group-administered tests. School districts as well as researchers often employ the use of group tests, mainly due to the expense and time constraints, but also because information from group tests can help educators make better instructional decisions for students and drive instructional and curriculum goals.

In order to accommodate the testing of students with disabilities, modifications in testing procedures are followed as written in the students Individualized Education Program (IEP) or Section 504 Plan. Such accommodations may include: administration in a small group; reading directions and test questions aloud; alternative answer procedures (pointing, oral); repetition of directions; frequent breaks; and extended time.

Lohman and Hagan (2001) define a testing accommodation as a change in the procedures for administering the assessment and that an accommodation is intended to neutralize, as much as possible, the effect of the student's disability on the assessment process. Accommodations should not change the kind of reasoning ability or skill being measured but should change how the ability or skill is measured. If chosen appropriately, an accommodation should provide neither too much nor too little help to the student who receives it.

Cognitive Abilities Test (CogAT)

In terms of theory and research basis of the CogAT, the construction of the instrument was based on concepts drawn from recent revisions of two classic models of human abilities: Vernon's (1961) hierarchical abilities model and Cattell's (1987) fluid-crystallized abilities model. Both of these are factor-analytic and hierarchical models in the sense that they posit several strata of factors ranging from very general to very specific. The two models differ in the number of general factors provided. Vernon's model describes one general reasoning factor called *g* that dominates all other factors. Cattell's model provides for two general reasoning factors, fluid-analytic reasoning (usually designated G*f*) and crystallized reasoning (usually designated G*c*), which dominate the other factors in the model. Of the two models, Cattell's is generally considered more comprehensive and complete, especially in its more recent revisions.

Vernon's model provides for two major group factors. One of these is called *v: ed*, or *verbal*: *educational*. It dominates the more specific factors such as logical reasoning, inductive reasoning, fluency, number facility, and vocabulary. The other major group factor is called *k: m*, or *spatial*: *mechanical*. It dominates the more specific factors such as spatial ability, mechanical information, and psychomotor coordination. Both major group factors are dominated by the general reasoning factor, *g*. The major concept taken from Vernon's model and applied to the construction of the CogAT is the three-stratum structure with a general reasoning factor, *g*, dominating all others and with major group factors dominating the more specific factors.

In 1943, Cattell first proposed his theory of fluid-crystallized abilities, generally called the G*f* - G*c* theory, but he did not fully develop and explicate it until 1963. In recent years, much of the research on Cattell's theory has been conducted by Horn (1985), a colleague who studied with Cattell, and the theory today is generally referred to as the Cattell-Horn theory. Cattell defined fluid-analytic abilities *(*G*f)* as general abilities that enable the individual to process mental information accurately and appropriately across a wide range of cognitive tasks. Cattell indicated that G*f* represented an individual's basic capacity to acquire knowledge and that cognitive tasks that require inductive, deductive or spatial reasoning were excellent measures of G*f*. He defined crystallized abilities (G*c*) as those that an individual acquired (partly on the basis of her or his level of G*f*) through acculturation, which includes both general experience and formal schooling. Cattell would include language development, vocabulary or lexical knowledge, verbal comprehension, and number facility under the rubric of crystallized abilities. Cattell stated that fluid abilities are much more likely to influence crystallized abilities than the reverse. Horn, on the other hand, emphasized that both G*c* and G*f* were developed through education and experience. Cattell and (later) Horn demonstrated that test tasks using geometric and other nonverbal symbols were excellent measures of G*f* and were less affected by acculturation than verbal and quantitative tasks. The authors of CogAT used Cattell's theory much more extensively than Vernon's to construct CogAT. The provision of three separate indexes, one of which is a nonverbal battery, is based on the research of Cattell and Horn. The concept of fluid-crystallized abilities also has been used to guide the selection of tasks for the test.

Although Cattell did not provide for a single, overall factor, *g*, Gustafsson (1988) demonstrated that G*f* and G*c* have substantial positive correlations, indicating that a higher factor dominates both of them. His analyses demonstrate that the models proposed by Vernon and Cattell are basically similar, although they use different terminology and provide for a different number of major group factors.

In the most comprehensive review to date, Carroll (1993) reanalyzed the work of Horn and Cattell, yielding conclusions in general agreement with those of other researchers in the field (see, for example, Gustafsson & Undheim, 1996). Carroll's first important finding is that human abilities are organized hierarchically. This means that some cognitive competencies are more broadly useful than others. It also means theories that postulate either an independent set of abilities or just one general ability are fundamentally flawed. The hierarchy that Carroll proposes starts with *g*, general mental ability, at the topmost level. Eight broad group factors define the second level, originally identified by Cattell and Horn. These factors vary in their association with *g*. The closest is the factor Cattell termed G*f*, or general fluid ability. Other factors at this level include G*c* (verbal crystallized ability), G*v* (broad visual perception), G*s* (processing speed), G*sm* (short term acquisition and retrieval), G*lr* (long term storage and retrieval), G*a* (broad auditory perception), and G*q* (quantitative knowledge). Finally, a longer list of psychologically more transparent primary factors defines the third level.

Carroll's analyses show that *g*, the topmost factor in the hierarchy, is most highly correlated with the second-level factor, G*f*, general fluid ability. G*f* in turn, is virtually synonymous with the primary factor, *r*, or reasoning. Gustafsson (1988) claims that the three factors are identical, that is, $g = Gf = r$. Others would describe the relationship between *g* and G*f* as more of an approximation than an identity. In either case, however, we are left with the important insight that reasoning abilities are at the core of human cognitive competence.

Carroll's secondary factor analysis helped to identify major aspects of individual differences in reasoning. He has shown that the G*f* factor may be separated into three subfactors: (1) sequential reasoning – verbal, logical, or deductive reasoning; (2) quantitative reasoning – inductive or deductive reasoning with quantitative concepts; and (3) inductive reasoning – typically measured with figural tasks. The CogAT was designed to measure all three aspects of reasoning, with a focus on the g factor at the third stratum of the CHC theory and the stratum II fluid reasoning abilities that load most highly on g (Lohman & Hagan, 2002).

Purpose of Current Study

As noted, previous research has documented the correlations between general cognitive ability, CHC abilities and performance on standardized achievement measures. The purpose of the current study is to examine the relationship between cognitive ability and measures of achievement, specifically, standardized achievement scores on the NJ ASK and school grades. Specifically, four research questions will be addressed: (1) What is the relationship between cognitive ability as measured by the CogAT and academic achievement as measured by NJ ASK test scores in grades 3, 5, and 7? (2) What is the relationship between cognitive ability as measured by the CogAT in grade 3 and grade 7 grades? Additionally, what is the relationship between academic achievement as measured by NJ ASK test scores in grade 3 and grade 7 grades? (3) Which of the five cluster scores on the grade 3 CogAT and NJ ASK (CogAT Verbal, CogAT Quantitative, CogAT Nonverbal, NJ ASK Language Arts and NJ ASK Mathematics) best predicts achievement in each of the academic subject areas investigated? (4) Can cognitive ability as measured by the CogAT and academic achievement as measured by NJ ASK test scores in grade 3 predict special education classification in grade 7?

CHAPTER 3

METHOD

Participants

The participants in this study consisted of a subsample (N = 452) of the 604 seventh grade students enrolled in a large, suburban public school district during the 2007-2008 school year. The students were approximately equally distributed across gender. Participants were selected based on the availability of test scores from two group-administered tests: the New Jersey Assessment of Skills and Knowledge (NJ ASK) and the Cognitive Abilities Test (CogAT). The sample consisted of seventh grade students who had valid scores on both tests in grades 3 (during the 2004 school year), 5 (during the 2006 school year) and 7 (during the 2008 school year). Only those participants for whom all six scores were available (N = 452) were included in the analysis.

The target school is a Kindergarten through grade 8 (K-8) public school district located in a middle to upper middle class suburb in the northeastern part of the United States. Five percent of students receive free and/or reduced lunches. The racial makeup of the district is 86% White, 6% Asian, 5% Black, 2% Hispanic, 1% Pacific Islander and less than 1% Native American. Currently staffed by approximately 450 teachers, there are approximately 5,000 students enrolled in grades K-8.

Materials

Materials/Measures Needed for the Current Study

SPSS Statistical Software

This software program was used to conduct all the intended analyses. As a statistical instrument SPSS provides a powerful set of sophisticated univariate and multivariate analysis techniques to fit the inherent characteristics of data describing complex relationships.

Cognitive Abilities Test, Form 6 (CogAT)

The Cognitive Abilities Test (CogAT) is a group-administered test designed to measure general abstract reasoning abilities and a student's capacity to apply these abilities to verbal, quantitative, and nonverbal cognitive tasks. This information can help educators make better instructional decisions for individuals and groups. The primary uses of CogAT scores are (1) to guide efforts to adapt instruction to the needs and abilities of students, (2) to provide an alternative measure of cognitive development, and (3) to identify students whose predicted levels of achievement differ markedly from their observed levels of achievements (Lohman & Hagan, 2001).

Form 6 of CogAT consists of three separate indexes that measure verbal, quantitative, and nonverbal reasoning. In addition to these index scores, it also yields a composite score that is based on the student's performance on all three indexes. The composite score indicates the variety and strength of the student's own cognitive resources for learning and the effectiveness with which individuals can use these resources to accomplish a wide range of cognitive tasks (Lohman & Hagen, 2001).

CogAT scores are reported in the form of Universal Scale Scores (USS) and Standard Age Scores (SAS). The USS is a normalized standard score and is the fundamental CogAT scale. It spans all levels of CogAT and is the means by which a raw score on one level of the test can be related to a raw score on any other test level. USS are considered developmental scores and were developed primarily to serve technical and research purposes. Lohman and Hagen (2001) do not recommend the use of USS scores for determining patterns of cognitive development because the mean USS scores on the three batteries are not identical. The Standard Age Scores (SAS) scale for each separate battery was developed using smoothed cumulative frequency distributions of USS scores of students at common age levels. The SAS scale ranges from a low of 50 to a high of 150. In all age groups, the mean SAS is 100, and the standard deviation is 16 SAS units. The SAS permit one to compare the rate and level of cognitive development of an individual to other students in the same group. When using SAS to determine the level of a student's cognitive skills, it is useful to think of them as clusters of scores that can be categorized as follows: SAS 50 – 72 Very Low; SAS 73 – 88 Below Average; SAS 89 – 111 Average; SAS 112 – 127 Above Average; SAS 128 – 150 Very High (Lohman & Hagen, 2001).

The Multilevel Edition of the CogAT Form 6 (Levels A to H) is administered to students in grades 3 through 12. Test Level A is typically administered to students in grade 3, with test level B given to students in grade 4, and so on. The Multilevel Verbal (65 items), Quantitative (60 items), and Nonverbal batteries (65 items) each contain three subtests that use different item formats. The directions for each subtest are read aloud by the test administrator while the students read them silently. Students read and respond to the items on their own and mark their answers on a separate answer sheet. Each subtest in a battery has a time limit that must be strictly observed. These limits were initially set to provide enough time for the majority of students to attempt to answer all of the items. All of the subtests except Quantitative Relations use multiple-choice items with five answer choices. The Quantitative Relations items present only three answer choices, and the choices follow the same pattern for each item. Table 3-1 shows the number of items and working time for each subtest in the Multilevel Edition (Lohman & Hagen, 2001).

Table 3-1. Number of Items and Working Time for Each Subtest on the CogAT

	Number of Items	Working Minutes (minutes)
Verbal Battery	65	30
Test 1: Verbal Classification	20	10
Test 2: Sentence Completion	20	10
Test 3: Verbal Analogies	25	10
Quantitative Battery	60	30
Test 4: Quantitative Relations	25	8
Test 5: Number Series	20	10
Test 6: Equation Building	15	12
Nonverbal Battery	65	30
Test 7: Figure Classification	25	10
Test 8: Figure Analogies	25	10
Test 9: Figure Analysis	15	10
Total	190	90

There are three subtests on the Verbal Battery: Verbal Classification, Sentence Completion and Verbal Analogies. One each subtest, each item presents five answer choices. On the Verbal Classification test, three words make up the stimulus. Students must discover the conceptual link among the stimulus words and then select from the answer choices the word that belongs with the stimulus words. For the Sentence Completion test, each item presents a sentence with one word missing. Each item on the Verbal Analogies test presents a verbal analogy in the form A→B: C →__. Students must figure out the relationship between words A and B and then select the word from among the answer choices that relates to word C in such a way to correctly complete the analogy (Lohman & Hagen, 2001).

The three subtests on the Quantitative Battery are Quantitative Relations, Number Series, and Equation Building. Depending on the level of the test, the Quantitative Relations test has either one or two item types. There are always three answer choices on this test (a form of “greater than,” “less than,” and “equal to”) and they are always presented in the same order. One type of item presents a pair of quantities (e.g., 3 dollars and 12 quarters). Students must judge whether the first quantity is greater than, less than, or equal to the second quantity. The second type notates relationships among abstract symbols and requires students to use the symbolic information to figure out the relationship between two of the variables. (For example, given $x = y$ and $z < y$, which is correct: x is greater than z, x is less than z, or x is equal to z?). On the Number Series test, students must figure out the rule underlying the progression in the presented series and select the next number in the series from among the five answer choices. Each item on the Equation Building test presents two or more numerals and one or more operational symbols. Students must combine the given stimuli to make an equation whose solution is given as one of the five answer choices (Lohman & Hagen, 2001).

The three subtests on the Nonverbal Battery are Figure Classification, Figure Analogies, and Figure Analysis. All of the items on this battery use figures, designs, or geometric shapes as stimuli, and all present five answer choices. Each item on the Figure Classification test presents a set of three stimulus figures. Students must determine the underlying characteristic that is common to the three figures and then select from among the answer choices the figure that also belongs with them. For the Figure Analogies test, three figures are presented in an analogy (A→B: C→_). Students must determine the relationship between Figure A and Figure B and then select from among the answer choices the figure that correctly completes the analogy. The difficulty among these items varies by the number and subtlety of the transformations used in the A→B relationship. Each item in the Figure Analysis test uses a series of diagrams to show how a square piece of paper is folded and where holes are punched in it. Students must select the answer choice that shows how the paper will look when it is unfolded. None of the subtests on this battery require reading or the use of English. None requires prior formal conceptual or factual knowledge. All of the information needed to answer an item correctly is embedded in the item itself (Lohman & Hagen, 2001).

The CogAT was normed on 149,798 students in grades K-12 during the spring of 2000. The sample was stratified on demographic variables such as geographic region, socioeconomic status, and ethnic groups as well as a balance between public and private schools. Sampling for the standardization produced a national probability sample representative of students nationwide (Lohman & Hagan, 2002).

In the present study, a different level of CogAT was administered at each grade level (Level A = Grade 3; Level C = Grade 5; Level E = 7). As a result of vertical scaling, the series yield comparable scores that can describe the longitudinal development of an individual's general verbal, quantitative, and nonverbal reasoning abilities as s/he progresses through school. Items on each test form were developed through an extensive tryout process that included screening for difficulty, discrimination, and differential item functioning (DIF). Items within each battery were then independently scaled to create a unidimentional, cross-grade scale for each battery. Kuder-Richardson Formula 20 (K-R

20), or coefficient alpha reliabilities, average .95 for the Verbal Battery, .94 for the Quantitative Battery, and .95 for the Nonverbal Battery for both fall and spring administrations. CogAT Composite scores are highly reliable. The three-battery Composite reliabilities average .98 for both the fall and spring administrations. Taken from the *Cognitive Abilities Test, Form 6 (CogAT)* research manual (2002), the reliability and average standard error of measurement of the CogAT USS for test batteries and composite scores used in this study are presented in Table 3-2 by grade.

Table 3-2. CogAT6 Universal Scale Score Reliability Coefficients for Spring Administration by Grade

Grade	Test Battery or Composite	Mean	SD	SEM	K-R 20
3	Verbal (V)	175.4	21.80	4.7	.953
	Quantitative (Q)	178.0	22.65	5.4	.943
	Nonverbal (N)	185.8	22.49	4.6	.958
	Composite (V+Q+N)	180.2	19.96	2.8	.980
5	Verbal (V)	192.0	20.70	4.6	.951
	Quantitative (Q)	193.0	20.83	5.2	.937
	Nonverbal (N)	196.6	20.89	4.4	.955
	Composite (V+Q+N)	194.2	18.67	2.7	.978
7	Verbal (V)	205.3	19.84	4.6	.946
	Quantitative (Q)	203.2	19.80	5.1	.934
	Nonverbal (N)	203.9	19.59	4.4	.950
	Composite (V+Q+N)	203.6	17.71	2.7	.977

CogAT was co-normed with the Iowa Tests of Basic Skills (ITBS) on 149,798 students in grades K-8 and with the Iowa Test of Educational Development (ITED) on 30,740 students in grade 9 – 12. The concurrent prediction of achievement was quite high. Average correlations with the ITBS Composite score were .83 for the Verbal

Battery, .78 for the Quantitative Battery, .71 for the Nonverbal Battery, and .86 for the CogAT Composite. Differential validity was also observed for the CogAT Verbal score (which showed highest correlations with reading and language tests on the achievement batteries) and for the CogAT Quantitative score (which showed highest correlations with mathematics test on the achievement batteries). The Nonverbal Battery showed the lowest correlations with all achievement scores. Put differently, the Nonverbal Battery provides the most independent information about the student's cognitive development (Lohman & Hagen, 2002).

In regard to internal consistency, results for the Multilevel Edition were remarkably consistent across all eight levels. Table 3-3 shows the factor loadings by test level (and grade level) for the hierarchical factor analyses of the multilevel edition (Lohman & Hagen, 2002).

Concurrent Validity. Lohman (2003a) investigated the concurrent validity of the Wechsler Intelligence Scale for Children – Third Edition (WISC-III; Wechsler, 1991) and the CogAT. The standard battery of WISC-III and level D of CogAT were administered to 91 sixth grade students. General ability as estimated by the Full Scale score in the WISC-III and by the Composite SAS score on CogAT correlated at least $r = .79$. Latent general factors on the two batteries correlated at least $r = .97$. The authors concluded that the general intellectual ability factor measured by the CogAT was the same general ability factor that was measured by the WISC-III.

Patterns of correlations between the CogAT Verbal, Quantitative, and Nonverbal batteries and the WISC-III Verbal and Performance scores revealed the following correlations: the CogAT Verbal SAS correlated highest with WISC-III Verbal ($r = .75$).

Table 3-3. Factor Loadings by Test Level for the Hierarchical Factor Analysis of the CogAT Multilevel Edition

	G	V’	Q’	N’
Level A (Grade 3)				
Verbal Classification	.62	.33	.00	.00
Sentence Completion	.75	.40	.00	.00
Verbal Analogies	.80	.42	.00	.00
Quantitative Relations	.80	.00	.15	.00
Number Series	.81	.00	.16	.00
Equation Building	.77	.00	.15	.00
Figure Classification	.75	.00	.00	.32
Figure Analogies	.82	.00	.00	.35
Figure Analysis	.67	.00	.00	.28
Level C (Grade 5)				
Verbal Classification	.60	.36	.00	.00
Sentence Completion	.74	.44	.00	.00
Verbal Analogies	.78	.47	.00	.00
Quantitative Relations	.82	.00	.09	.00
Number Series	.81	.00	.08	.00
Equation Building	.78	.00	.08	.00
Figure Classification	.74	.00	.00	.34
Figure Analogies	.82	.00	.00	.38
Figure Analysis	.67	.00	.00	.31
Level E (Grade 7)				
Verbal Classification	.72	.44	.00	.00
Sentence Completion	.74	.45	.00	.00
Verbal Analogies	.78	.48	.00	.00
Quantitative Relations	.82	.00	.20	.00
Number Series	.83	.00	.20	.00
Equation Building	.75	.00	.18	.00
Figure Classification	.75	.00	.00	.34
Figure Analogies	.82	.00	.00	.37
Figure Analysis	.71	.00	.00	.32

CogAT Nonverbal SAS correlated about equally with both WISC-III Verbal (r = .59) and

WISC-III Performance (*r* = .57) scale scores. The CogAT Quantitative SAS also correlated *r* = .58 with the WISC-III Verbal but only *r* = .42 with the WISC-III Performance scale scores.

Correlations for latent variables showed that the two verbal batteries were most similar (*r* = .87). Although the latent factors for the WISC-III Performance Scale and the CogAT Nonverbal Battery correlated only *r* = .64, the other two CogAT batteries showed substantial correlations with the WISC-III Performance scale as well. Based on these data, the authors conclude that the WISC-III Performance scale cannot be equated with the CogAT Nonverbal Battery. Although there was a clear association between the two verbal batteries and a weaker association between the WISC-III Performance and CogAT Nonverbal batteries the overriding factor seems to be the presence of a common general factor in both test batteries.

Lohman (2003b) also examined the concurrent validity of the Woodcock-Johnson III (WJ-III; Woodcock, McGrew; & Mather, 2001) and the CogAT. A total of 178 students in grades 2, 5, and 9 were administered 9 subtests from the WJ-III standard battery plus four additional subtests (Planning, Analysis-Synthesis, Applied Problems, and Quantitative Concepts) and the appropriate level of the CogAT. Interbattery confirmatory factor analyses showed that the general factors on the two batteries correlated *r* = .82. Correlations between broad-group clusters on the WJ-III and battery-level scores on the CogAT generally supported the construct interpretations of each. For example, the CogAT Verbal correlated most highly with the WJ-III Verbal Ability score (*r* = .62); the CogAT Quantitative correlated most highly with the WJ-III Math

Reasoning (r = .58); and CogAT Nonverbal correlated most highly with WJ-III Fluid Reasoning (r = .55).

Correlations between broad-group clusters on the WJ-III and battery-level scores on the CogAT also suggested important differences in the abilities measured by both batteries. For example, the CogAT Verbal factor had its highest correlations with G*c*, followed closely by G*q*. Further, both the CogAT Quantitative and Nonverbal factors showed their highest correlations with G*q*. In general, the CogAT Verbal, Quantitative, and Nonverbal scores correlated higher with the WJ-III General Intellectual Ability (GIA) cluster than with any of the more specific WJ-III clusters. The authors conclude that all three of the CogAT batteries best measure what is captured by the GIA cluster and only secondarily what is measured by the five more specific clusters.

Theoretical Schema for the CogAT

The CogAT appraises general abstract reasoning abilities and the student's capacity to apply these abilities to verbal, quantitative, and nonverbal cognitive tasks. The test is divided into three batteries, each representing one of these three areas. All three batteries appraise inductive and deductive reasoning, which Cattell would classify as fluid-analytic abilities. The Verbal and Quantitative batteries appraise additional cognitive abilities that are specifically related to those batteries, such as verbal abilities on the Verbal Battery. Cattell would classify these as crystallized abilities. The Composite score for all three batteries is a measure of overall cognitive ability, termed *g*.

New Jersey Assessment of Skills and Knowledge (NJ ASK)

The New Jersey Assessment of Skills and Knowledge (NJ ASK) was designed to give an early indication of the progress students make in mastering the knowledge and skills described in the New Jersey Core Curriculum Content Standards (CCCS) (New Jersey Department of Education, 2004; New Jersey Department of Education, 2006). The test also fulfills the requirements under the federal No Child Left Behind (NCLB) Act. The results are to be used by school districts to identify strengths and weaknesses in their educational programs. This process is expected to lead to improved instruction and better alignment with the CCCS. The results are also used, along with other indicators of student progress, to identify those students who may need instructional support in any of the content areas. This support, which could be in the form of individual or programmatic intervention, would be a means to address any identified knowledge or skill gaps.

The NJ ASK consists of two content areas: Language Arts Literacy and Mathematics. Scores at grades 3 – 7 are reported as scale scores with a range of 100 to 300 in each of the content areas. The scale scores are intended to be comparable across forms within a grade. NJ ASK scale scores are not comparable across subjects (e.g., LAL and Math) or grades (e.g., grade 3 and grade 4). The scores range from 100-199 (Partially Proficient), 200-249 (Proficient), and 250-300 (Advanced Proficient). The scores of students who are included in the Partially Proficient level are considered to be below the state minimum of proficiency and those students may be most in need of instructional support (NJ DOE, 2004; NJ DOE, 2006).

The NJ ASK administered to students in grades 5, 6, and 7 is designed to measure the same CCCS as the NJ ASK administered to students in grades 3 and 4. The Language Arts Literacy section of each test measures students' achievement in reading and writing. Students read passages selected from published books, newspapers, magazines, and everyday text and respond to related multiple-choice questions, open-ended questions and a writing task. The Language Arts Literacy assessment measures knowledge and skills in the following clusters in grades 3 – 7 (a cluster is a group of related test questions on a single topic): Writing and Reading (Working with Text and Analyzing Text). In grades 3 and 4, the Writing cluster is broken down into two clusters: Writing about Pictures and Writing about Poems.

The Mathematics section of each test measures students' ability to solve problems by applying mathematical concepts. The NJ ASK assesses four CCCS in Mathematics in grades 3 – 7: (1) Number Sense and Numerical Operations (2) Geometry and Measurement (3) Patterns and Algebra and (4) Data Analysis, Probability, and Discrete Mathematics. A process cluster, Problem Solving, is also provided on score reports. The process cluster refers to test questions that measure mathematical problem-solving ability. Each test question on the Mathematics assessment measures one content cluster and may contribute to the process cluster. For grades 5 – 7, each content cluster in Mathematics contains one of three open-ended items. For grades 3 and 4, each cluster contains one open-ended item.

The NJ ASK 3 Language Arts Literacy and Mathematics tests were administered to 104,962 total students in grade 3 in spring of 2004. Of the 103,414 grade 3 students with valid scale scores in Language Arts Literacy, 20.7% scored in the Partially Proficient range; 75.6% scored in Proficient range and 3.8% scored in Advanced Proficient range. The mean scale score in the Language Arts Literacy content area was 215.5. Of the 103,559 grade 3 students with valid scores in Mathematics, 23.4% scored in Partially Proficient range; 53.8% scored in Proficient range and 22.8% scored in Advanced Proficient range. The mean scale score in the Mathematics content area was

222.2. For grade 3, the internal reliability coefficient is .84 and .86 for Language Arts Literacy and Mathematics, respectively. The average standard error of measurement is 2.28 and 2.45 for Language Arts Literacy and Mathematics, respectively. Reliability and standard errors of measurement (SEM) for content areas and clusters for grade 3 are reported in table 3-4 (NJ DOE, 2004).

Table 3-4. Reliability Estimates and Standard Errors of Measurement (SEM) for Content Areas and Clusters for the NJ ASK 3

NJ ASK Test Section	Reliability	Raw Score SEM
Language Arts Literacy	0.84	2.28
Reading	0.80	1.66
Writing	0.72	1.37
Working with Text	0.70	1.14
Analyzing Text	0.65	1.21
Mathematics	0.86	2.45
Number Sense and Numerical Operations	0.71	1.10
Geometry and Measurement	0.44	1.37
Patterns and Algebra	0.57	1.28
Data analysis, Probability and Discrete Math	0.60	1.33
Problem Solving	0.83	2.23

The NJ ASK 5 Language Arts Literacy and Mathematics tests were administered to 104,418 total students in grade 5 in spring of 2006. Of the grade 5 students with valid scale scores in Language Arts Literacy, 14.1% scored in the Partially Proficient range; 76.6% scored in Proficient range and 9.3% scored in Advanced Proficient range. The mean scale score in the Language Arts Literacy content area was 220.1. Of the grade 5 students with valid scores in Mathematics, 18.3% scored in Partially Proficient range; 54.2% scored in Proficient range and 27.5% scored in Advanced Proficient range. The mean scale score in the Mathematics content area was 227.6. For grade 5, the internal reliability coefficient is .85 and .84 for Language Arts Literacy and Mathematics, respectively. The average standard error of measurement is 2.28 and 2.91 for Language Arts Literacy and Mathematics, respectively. Reliability and standard errors of measurement (SEM) for content areas and clusters for grade 5 are reported in table 3-5 (NJ DOE, 2006).

Table 3-5. Reliability Estimates and Standard Errors of Measurement (SEM) for Content Areas and Clusters for the NJ ASK 5

NJ ASK Test Section	Reliability	Raw Score SEM
Language Arts Literacy	0.85	2.28
Reading	0.85	2.15
Writing		
Working with Text	0.74	1.59
Analyzing Text	0.73	1.46
Mathematics	0.84	2.91
Number Sense and Numerical Operations	0.58	1.48
Geometry and Measurement	0.46	1.23
Patterns and Algebra	0.56	1.47
Data analysis, Probability and Discrete Math	0.62	1.60
Problem Solving	0.83	2.86

The NJ ASK 7 Language Arts Literacy and Mathematics tests were administered to 105,518 total students in grade 7 in spring of 2008. Of the 103,995 grade 7 students with valid scale scores in Language Arts Literacy, 29.6% scored in the Partially Proficient range; 55.7% scored in Proficient range and 14.7% scored in Advanced Proficient range. The mean scale score in the Language Arts Literacy content area was 215.4. Of the 104,393 grade 7 students with valid scores in Mathematics, 35.7% scored in

Partially Proficient range; 44.5% scored in Proficient range and 19.8% scored in Advanced Proficient range. The mean scale score in the Mathematics content area was 213.3. For grade 7, the internal reliability coefficient is .89 and .91 for Language Arts Literacy and Mathematics, respectively. The average standard error of measurement is 3.58 and 3.26 for Language Arts Literacy and Mathematics, respectively (J. Liang, personal communication, April 14, 2009).

Procedure

The proposal for this research was sent to the Institutional Review Board (IRB) for approval on February 5, 2008. Exemption from full committee IRB review was granted on February 15, 2008, as the current study proposed to analyze existing data. Archived school records for the subjects were examined for their scores on the CogAT and NJ ASK in grades 3 (during the 2004 school year), 5 (during the 2006 school year), and 7 (during the 2008 school year). Grades from the 2007-2008 school year in academic subject areas were collected from a file review at the school. Information from school records was collected on coded forms in an effort to maintain confidentiality.

The researcher created a database of scores that included the three indexes of the CogAT: verbal, quantitative, and nonverbal reasoning. The composite score was used as an overall measure of *g*. Scores from grades 3, 5, and 7 were collected. NJ ASK scores from grades 3, 5 and 7 in the two content areas, Language Arts Literacy and Math were also collected. Proficiency status was designated by one of three rankings: Partially Proficient, Proficient and Advanced Proficient.

Grades in reading, writing, math, social studies and science were collected from report cards from the last trimester (or third grading period) of seventh grade during the

2007-2008 school year. These subjects were chosen as the academic subjects under review in this study because they are the core classes that all students are required to take. The target school converts number grades to letter grades as follows: A+ = 97-100; A = 93-96; A- = 90-92; B+ = 87-89; B = 83-86; B- = 80-82; C+ = 77-79; C = 73-76; C- = 70-72; D+ = 67-69; D = 63-66; D- = 60-62; F = 59 and below. Letter grades were converted to a point value system and entered into the database. Specifically, A+ = 4.3; A = 4.0; A- = 3.7; B+ = 3.3; B = 3.0; B- = 2.7; C+ = 2.3; C = 2.0; C- = 1.7; D+ = 1.3; D = 1.0; D- = 0.7; F = 0. Other general indicators, such as gender, ethnicity, and special education classification were also collected.

Design and Analysis

This study investigated archival data of a select subsample (N = 452) of the 604 seventh grade students enrolled in a large, suburban public school district during the 2007-2008 school year. The current study examined the following questions: (1) What is the relationship between cognitive ability as measured by the CogAT and academic achievement as measured by NJ ASK test scores in grades 3, 5, and 7? (2) What is the relationship between cognitive ability as measured by the CogAT in grade 3 and grade 7 grades? Additionally, what is the relationship between academic achievement as measured by NJ ASK test scores in grade 3 and grade 7 grades? (3) Which of the five cluster scores on the grade 3 CogAT and NJ ASK (CogAT Verbal, CogAT Quantitative, CogAT Nonverbal, NJ ASK Language Arts and NJ ASK Mathematics) best predicts achievement in each of the academic subject areas investigated? (4) Can cognitive ability as measured by the CogAT and academic achievement as measured by NJ ASK test scores in grade 3 predict special education classification in grade 7?

In order to answer the first question regarding the relationship between cognitive abilities and academic achievement, Pearson correlations were computed between CogAT and NJ ASK scores by grade. In order to determine if the correlations between the NJ ASK and CogAT scores were statistically different from each other, correlation coefficients were converted to Fisher's z scores (Mertler & Vannatta, 2005) and the significance of the difference between the correlations was tested. To answer the second question, separate Pearson correlation analyses were run. The first correlation analysis examined the relationship between CogAT scores in grade 3 and grade 7 grades in each academic subject area. The second correlation analysis examined the relationship between NJ ASK scores in grade 3 and grade 7 grades in each academic subject area. In order to answer the third question, a full scale multiple regression investigated which CogAT and/or NJ ASK score(s) in grade 3 had the greatest weight in predicting achievement in each of the academic subject areas investigated. In order to answer the final research question, a separate-sample t-test was computed comparing the mean scores on the NJ ASK Language Arts, NJ ASK Mathematics, CogAT Verbal, CogAT Quantitative and CogAT Nonverbal scores for classified and non-classified students. Then, a logistic regression was computed for CogAT and NJ ASK scores and special education classification in order to determine of CogAT and NJ ASK test scores in grade 3 could predict special education classification in grade 7.

CHAPTER 4

RESULTS

The current study analyzed the relationships between cognitive ability and measures of achievement, specifically, standardized achievement scores on the New Jersey Assessment of Skills and Knowledge (NJ ASK) and school grades. Correlations, multiple regressions and t-tests were conducted in order to identify these relationships. Four research questions were addressed and the results are presented below.

As an initial analysis, frequency distributions were computed for the demographic variables. These data are contained in Table 4-1.

Cognitive Ability and Academic Achievement

The first research question asked: What is the relationship between cognitive ability as measured by the Cognitive Abilities Test (CogAT) and academic achievement as measured by NJ ASK test scores in grades 3, 5, and 7? Table 4-2 presents means and standard deviations for the NJ ASK Language Arts and Mathematics scores and the CogAT Verbal, Quantitative, Nonverbal and Composite scores by grade.

To answer the first research question, Pearson correlations were computed between CogAT and NJ ASK scores by grade. These correlations are presented in Tables 4-3, 4-4, and 4-5. In order to determine if the correlations between the NJ ASK and CogAT scores were statistically different from each other, correlation coefficients were converted to Fisher's z scores (Mertler & Vannatta, 2005) and the significance of the difference between the correlations was tested. These scores are presented in Tables 4-6, 4-7, 4-8 by grade.

Table 4-1. Descriptive Statistics for Demographic Variables

	N	Percent
Gender		
Female	223	50.7
Male	229	49.3
Ethnicity		
White	400	88.5
Asian	24	5.3
Black	17	3.8
Multi-racial	6	1.3
Hispanic	5	1.1
Economically Disadvantaged		
No	439	96.5
Yes	16	3.5
Special Education Classification		
No	379	83.8
Yes	73	16.2

While the NJ ASK scores and CogAT scores are significantly correlated with each other at each grade level, only some scores were statistically different from each other. In grades 3, 5, and 7, the CogAT Verbal score showed a significantly higher correlation with the NJ ASK Language Arts score, when compared to the CogAT Quantitative and Nonverbal scores. Only in grade 7 did the CogAT Quantitative

Table 4-2. Means and Standard Deviations of NJ ASK and CogAT Scores by Grade

	Range	Mean	Standard Deviation
Grade 3			
NJ ASK Language Arts	172-262	225.91	16.970
NJ ASK Mathematics	157-276	232.26	24.682
CogAT Verbal	75-149	106.57	12.880
CogAT Quantitative	74-149	110.83	13.369
CogAT Nonverbal	74-150	110.02	13.485
CogAT Composite	74-149	109.94	12.754
Grade 5			
NJ ASK Language Arts	174-300	226.14	17.684
NJ ASK Mathematics	154-300	236.02	28.131
CogAT Verbal	73-150	108.28	12.376
CogAT Quantitative	79-149	114.14	12.865
CogAT Nonverbal	72-150	113.54	13.023
CogAT Composite	76-150	113.09	12.559
Grade 7			
NJ ASK Language Arts	152-300	218.40	26.221
NJ ASK Mathematics	134-300	219.67	33.225
CogAT Verbal	77-150	107.04	12.084
CogAT Quantitative	71-150	110.31	14.012
CogAT Nonverbal	69-150	111.30	13.210

Table 4-2. (continued)

	Range	Mean	Standard Deviation
CogAT Composite	72-148	110.77	12.702

Note. N = 452.

Table 4-3. Pearson Correlations between CogAT and NJ ASK Scores Grade 3

	CogAT Verbal	CogAT Quantitative	CogAT Nonverbal
NJ ASK Language Arts	.548**	.404**	.346**
NJ ASK Mathematics	.581**	.595**	.594**

Note. N = 452.

** $p < 0.01$

Table 4-4. Pearson Correlations between CogAT and NJ ASK Scores Grade 5

	CogAT Verbal	CogAT Quantitative	CogAT Nonverbal
NJ ASK Language Arts	.588**	.415**	.471**
NJ ASK Mathematics	.597**	.572**	.599**

Note. N = 452.

** $p < 0.01$

Table 4-5. Pearson Correlations between CogAT and NJ ASK Scores Grade 7

	CogAT Verbal	CogAT Quantitative	CogAT Nonverbal
NJ ASK Language Arts	.725**	.524**	.497**
NJ ASK Mathematics	.642**	.751**	.673**

Note. N = 452.

** $p < 0.01$

Table 4-6. NJ ASK and CogAT Fisher's z Scores Grade 3

NJ ASK Language Arts	
CogAT Verbal > Quantitative	z = 2.88 *
CogAT Verbal > Nonverbal	z = 3.76 *
CogAT Quantitative > Nonverbal	z = 0.88
NJ ASK Mathematics	
CogAT Verbal < Quantitative	z = 0.46
CogAT Verbal < Nonverbal	z = 0.23
CogAT Quantitative > Nonverbal	z = 0.23

Note. In order to be significant, a value of 1.96 is needed.

* $p < .05$

Table 4-7. NJ ASK and CogAT Fisher's z Scores Grade 5

NJ ASK Language Arts	
CogAT Verbal > Quantitative	z = 3.47 *
CogAT Verbal > Nonverbal	z = 2.53 *
CogAT Quantitative < Nonverbal	z = 0.94
NJ ASK Mathematics	
CogAT Verbal > Quantitative	z = 0.69
CogAT Verbal < Nonverbal	z = 0.00
CogAT Quantitative < Nonverbal	z = 0.69

Note. In order to be significant, a value of 1.96 is needed.

* $p < .05$

score show significantly higher correlations with the NJ ASK Mathematics score, when compared to the CogAT Verbal and Nonverbal scores. Therefore, in answering the first research question, the CogAT Verbal score and the NJ ASK Language Arts score showed significant correlations at grades 3, 5, and 7. Only in grade 7 did the CogAT Quantitative score and the NJ ASK Mathematics show significant correlations.

Cognitive Ability and Grades

The second research question asked: What is the relationship between cognitive ability as measured by the CogAT in grade 3 and grade 7 grades? Table 4-9 presents the means and standard deviations of grade 7 grades. To answer the second research question, Pearson correlations were computed between CogAT scores in grade 3 and

Table 4-8. NJ ASK and CogAT Fisher's z Scores Grade 7

NJ ASK Language Arts	
CogAT Verbal > Quantitative	z = 5.32 *
CogAT Verbal > Nonverbal	z = 5.72 *
CogAT Quantitative > Nonverbal	z = 0.41
NJ ASK Mathematics	
CogAT Verbal < Quantitative	z = 3.24 *
CogAT Verbal < Nonverbal	z = 0.79
CogAT Quantitative > Nonverbal	z = 2.45 *

Note. In order to be significant, a value of 1.96 is needed.

* p <.05

Table 4-9. Means and Standard Deviations of Grade 7 Grades

	Range	Mean	Standard Deviation
Reading	0-4.3	3.04	.87
Writing	0-4.3	3.24	.71
Math	0-4.3	3.17	.87
Science	0-4.3	3.34	.85
Social Studies	0-4.3	3.35	.91

Note. N = 452.

Grades were based on a 4 point scale with A+ = 4.3, A = 4.0, A- = 3.7, etc.

grade 7 grades. These correlations are presented in Table 4-10. As evidenced in Table 4-10, all correlations were statistically significant.

Table 4-10. Pearson Correlations between CogAT Scores in Grade 3 and Grade 7 Grades

	Reading	Writing	Math	Science	Social Studies
CogAT Verbal	.325**	.287**	.317**	.372**	.350**
CogAT Quantitative	.285**	.215**	.284**	.354**	.300**
CogAT Nonverbal	.316**	.278**	.354**	.376**	.318**

Note. N = 452.

** $p < 0.01$

The CogAT Verbal and Quantitative scores showed the same pattern of correlations, from highest to lowest: science, social studies, reading, math and writing. The CogAT Nonverbal correlation pattern differed slightly, with science correlating the highest followed by math, social studies, reading and writing. Therefore, in answering second research question, third grade cognitive ability was significantly correlated with grade 7 grades in each academic area studied.

Academic Achievement and Grades

A follow-up to the second research question asked: What is the relationship between academic achievement as measured by the NJ ASK test scores in grade 3 and grade 7 grades? To answer this question, Pearson correlations were computed between

NJ ASK scores in grade 3 and grade 7 grades. These correlations are presented in Table 4-11.

Table 4-11. Pearson Correlations between NJ ASK Scores in Grade 3 and Grade 7 Grades

	Reading	Writing	Math	Science	Social Studies
NJ ASK Language Arts	.425**	.403**	.328**	.390**	.398**
NJ ASK Mathematics	.415**	.366**	.450**	.464**	.451**

Note. N = 452.

** $p < 0.01$

As shown in Table 4-11, all correlations were statistically significant. The NJ ASK Language Arts score correlated the highest with grades in reading, followed by writing, social studies, science and math. The NJ ASK Mathematics correlated the highest with grades in science, followed by social studies, math, reading, and writing. Therefore, in answering the second research question, third grade achievement scores were significantly correlated with grade 7 grades in each academic area studied.

Cognitive Ability, Academic Achievement, and Predicted Achievement

The third research question asked: Which of the five cluster scores on the grade 3 CogAT and NJ ASK (CogAT Verbal, CogAT Quantitative, CogAT Nonverbal, NJ ASK Language Arts and NJ ASK Mathematics) best predicts achievement in each of the academic subject areas investigated? The results of a full-scale multiple regression are presented in Table 4-12.

Table 4-12. Full Scale Multiple Regression for CogAT and NJ ASK Scores and Academic Grades – Beta Weights

	Reading	Writing	Math	Science	Social Studies
NJ ASK Language Arts	.271*	.294*	.075	.144*	.172*
NJ ASK Mathematics	.168*	.141*	.328*	.233*	.265*
CogAT Verbal	.020	.009	.030	.070	.067
CogAT Quantitative	-.007	-.076	-.049	.047	-.002
CogAT Nonverbal	.116*	.136*	.147*	.120*	.066
R	.476	.440	.469	.502	.482
R Square	.226	.194	.220	.252	.232
Adj. R Square	.217	.185	.211	.244	.223
Sig of R	.000	.000	.000	.000	.000

* $p < .05$

As evidenced in Table 4-12, the full scale multiple regressions indicated that the NJ ASK Language Arts score was the best predictor of achievement in reading and writing. Further, the NJ ASK Mathematics score was the best predictor of achievement in math, science and social studies. In combination, the best predictors of grade 7 reading, writing and science grades were the NJ ASK Language Arts, Mathematics and CogAT Nonverbal scores in grade 3. In regards to grade 7 math grades, the best predictors, in combination, were NJ ASK Mathematics and CogAT Nonverbal scores. Finally, the best predictors of grade 7 social studies grades were the NJ ASK Language Arts and Mathematics scores.

Cognitive Ability, Academic Achievement, and Special Education Classification

The fourth research question asked: Can cognitive ability as measured by the CogAT and academic achievement as measured by NJ ASK test scores in grade 3 predict special education classification in grade 7? Of the 452 study participants, 83.8% (N = 379) were not classified as special education students; 16.2% (N = 73) had a special education classification in seventh grade. Table 4-13 presents the means and standard deviations for NJ ASK Language Arts, Mathematics, CogAT Verbal, CogAT Quantitative and CogAT Nonverbal scores by special education classification. A separate-sample t–test was computed comparing the mean scores on the NJ ASK Language Arts, NJ ASK Mathematics, CogAT Verbal, CogAT Quantitative and CogAT Nonverbal scores for classified and non-classified students in order to determine if there was a significant difference in scores between students with and without a special education classification. The results of this analysis are also contained in Table 4-13.

As evidenced from Table 4-13, the means between students with and without a special education classification, for all scores, were significantly different. Stated differently, all variables differentiated between special education classification.

In order to determine if CogAT and NJ ASK test scores in grade 3 could predict special education classification in grade 7, a logistic regression was computed for CogAT

Table 4-13. Means, Standard Deviations and Comparisons of Means by t-tests for NJ ASK Language Arts, NJ ASK Mathematics, CogAT Verbal, CogAT Quantitative and CogAT Nonverbal Scores by Special Education Classification

	Special Education	Mean	Standard Deviation	t-test
NJ ASK Language Arts	No	229.23	14.791	10.57 **
	Yes	208.68	17.215	
NJ ASK Mathematics	No	236.91	21.490	10.10 **
	Yes	208.11	26.196	
CogAT Verbal	No	108.51	12.427	7.76 **
	Yes	96.49	10.274	
CogAT Quantitative	No	112.40	12.839	5.87 **
	Yes	102.71	13.206	
CogAT Nonverbal	No	111.19	12.957	4.27 **
	Yes	103.97	14.611	

** p <.001

and NJ ASK scores and special education classification. These results are reported in Table 4-14.

As evidenced in Table 4-14, when all variables were combined, only NJ ASK Language Arts, NJ ASK Mathematics and CogAT Verbal scores remained as significant. Stated differently, third grade NJ ASK Language Arts, NJ ASK Mathematics and CogAT Verbal scores were the best predictors of special education classification in grade 7, accounting for a combined 22% of the variance. Table 4-15 shows the prediction results.

Table 4-14. Logistic Regression for CogAT and NJ ASK Scores and Special Education Classification

	B	Wald Statistic	Significance
NJ ASK Language Arts	-.041	14.02	.000**
NJ ASK Mathematics	-.029	11.30	.001**
CogAT Verbal	-.045	6.82	.009**
CogAT Quantitative	-.006	0.14	.706
CogAT Nonverbal	.031	3.42	.064

** *p* <.01

Table 4-15. Prediction Results for Special Education Classification

Observed	Predicted Not Classified	Predicted Classified
Table 4-15 (continued).		
Not Classified	368	11
Classified	49	24

The prediction table shows that 368 of the 379 (97%) non-classified students were correctly predicted as non-classified students. In contrast, of the 73 students actually classified as special education in seventh grade, only 33% (N = 24) of those students were correctly predicted as special education students. Stated differently, the analysis was successful in predicting students who were not classified as special education, with 97%

of non-special education students predicted as such. However, the analysis was not successful in predicting special education classification, with only 33% of special education students predicted as such. In other words, the probability of correctly identifying a student as special education is low (33%), while the probability of correctly identifying a student as not special education is high (97%).

CHAPTER 5

DISCUSSION

This study was designed to investigate the relationship between cognitive ability and measures of achievement, specifically, standardized achievement scores on the New Jersey Assessment of Skills and Knowledge (NJ ASK) and school grades. The literature in this field indicated that cognitive ability correlates with standardized achievement measures of reading (e.g., Evans et al., 2002; Flanagan et al., 2002; McGrew, 1993) and mathematics (e.g., Floyd, Evans, & McGrew, 2003; McGrew & Flanagan, 1998). However, a paucity of research exists regarding the relationship between cognitive ability or achievement test results and non-standardized measures of academic achievement such as grades in particular subject areas. Furthermore, research has not investigated the relationship between cognitive scores on the CogAT, academic scores on the NJ ASK and special education classification. Therefore, this study intended to answer four research questions: (1) What is the relationship between cognitive ability as measured by the CogAT and academic achievement as measured by NJ ASK test scores in grades 3, 5, and 7? (2) What is the relationship between cognitive ability as measured by the CogAT in grade 3 and grade 7 grades? Additionally, what is the relationship between academic achievement as measured by NJ ASK test scores in grade 3 and grade 7 grades? (3) Which of the five cluster scores on the grade 3 CogAT and NJ ASK (CogAT Verbal, CogAT Quantitative, CogAT Nonverbal, NJ ASK Language Arts and NJ ASK Mathematics) best predicts achievement in each of the academic subject areas investigated? (4) Can cognitive ability as measured by the CogAT and academic achievement as measured by NJ ASK test scores in grade 3 predict special education classification in grade 7?

(1) *Cognitive Ability and Academic Achievement*

Pearson correlations showed that CogAT scores and NJ ASK scores were significantly correlated with each other at each grade level. This is consistent with previous research demonstrating that tests of intelligence represent the best predictor of achievement in relation to other indices, with correlations between IQ scores and achievement test scores averaging about r = .40 to r = .50 (Sternberg, Grigorenko, & Bundy, 2001). However, upon further analysis, only some scores were statistically different from each other when the difference between the correlations was tested. In grades 3, 5, and 7, the CogAT Verbal score showed a significantly higher correlation with the NJ ASK Language Arts score, when compared to the CogAT Quantitative and Nonverbal scores. The results support the author's hypothesis as well as current research, as cognitive measures of crystallized knowledge (G*c*) (also termed 'verbal reasoning') have shown differential validity with standardized measures of reading (e.g., Evans et al., 2002; Flanagan et al., 2002; Lohman & Hagen, 2002; McGrew, 1993) and writing (McGrew and Knopik, 1993). This finding fits with the knowledge we have regarding what crystallized ability measures: This ability measures a person's use of verbal skills and their understanding of language development that is generally learned through classroom and cultural experiences. Since the Language Arts section of the NJ ASK focuses on word meanings and analogies, it is logical that CogAT Verbal scores loaded on the NJ ASK Language Arts scores more than the other CogAT abilities. Both abilities stem largely from the acquisition of declarative and procedural knowledge. Further, one of the narrow abilities under crystallized intelligence is defined as Lexical Knowledge, which is related to vocabulary development (McGrew & Flanagan, 1998). It also seems logical that prior exposure to the culture and language of the test should be of benefit to the test taker when attempting to garner understanding from a reading passage or define vocabulary (Vanderwood et al. 2002).

Only in grade 7 did the CogAT Quantitative score show significantly higher correlations with the NJ ASK Mathematics score, when compared to the CogAT Verbal and Nonverbal scores. While this finding only occurred in grade 7, it is as expected, as cognitive measures of fluid reasoning (G*f*) (e.g., Floyd, Evans & McGrew, 2003; McGrew & Flanagan, 1998) and more specifically, quantitative reasoning, have shown differential validity with standardized measures of mathematics (Lohman, 2003b). While inductive and deductive reasoning are generally considered to be the hallmark narrow-ability indicators of fluid reasoning, G*f* also subsumes more specific types of reasoning, most notably quantitative reasoning (RQ). In fact, Carroll's (1993) analysis and comprehensive review of human cognitive abilities indicated that quantitative reasoning abilities could be included under the stratum II ability, Fluid Intelligence. Quantitative reasoning is directly related to formal instruction and classroom related experiences. The correlation between the CogAT Quantitative and the NJ ASK Mathematics score supports Flanagan et al.'s (2002) explanation that RQ can be seen as a measure of learning, and in fact, subtests that measure mathematics reasoning on a number of achievement tests appear to directly measure RQ. At the very least, it represents a good example of one of those abilities that is often measured by both cognitive and achievement tests.

(2) *Cognitive Ability and Grades*

Pearson correlations found that significant relationships existed between CogAT scores in grade 3 and grade 7 grades in all academic subject areas investigated. The CogAT Verbal and Quantitative scores showed the same pattern of correlations, from highest to lowest: (1) science, (2) social studies, (3) reading, (4) math and (5) writing. While the order of correlations was not exactly as predicted for the CogAT Verbal score, it is understandable that science and social studies grades showed the highest correlations, since knowledge of the principles of science and social studies rely heavily on learned, factual information as well as the breadth and depth of a person's acquired knowledge base (i.e., crystallized knowledge or verbal reasoning ability). Further, one of the narrow abilities subsumed under crystallized knowledge is defined as General Science Information, which is one's range of scientific knowledge. Also, since crystallized knowledge subsumes narrow cognitive abilities associated with listening and speaking, the relations between the CogAT Verbal score and science and social studies grades may be due to the influence of language-based cognitive processes on performance (McGrew & Flanagan, 1998). A further explanation for this relationship is connected to the curriculum, instruction and assessment methods within the district of study. As outlined in the New Jersey social studies and science core curriculum content standards, teaching and learning activities heavily rely on skills in the areas of listening, speaking, reading, and written expression, all of which are consumed under the broad area of crystallized knowledge.

It is surprising that the CogAT Verbal score did not show stronger correlations with grades in reading and writing, particularly given the present study's results indicating correlations between the CogAT Verbal score and the NJ ASK Language Arts score, since both the NJ ASK Language Arts and grades in reading and writing are measures of reading and writing achievement. In addition, literature has linked crystallized ability to measures of reading (e.g., Evans et al., 2002; Flanagan et al., 2002; Lohman & Hagen, 2002; McGrew, 1993) and writing (McGrew & Knopik, 1993). This underscores an important point that successful school learning depends on many personal characteristics other than cognitive ability, such as persistence, interest in school and willingness to study. The encouragement for academic achievement that is received from peers, family and teachers may also be important, together with more general cultural factors (Neisser et al., 1996).

The order of correlations was not exactly as predicted for the CogAT Quantitative score. While the CogAT Quantitative score best predicted grades in science, it is surprising that math grades were among the weaker correlations, particularly since research has shown that subtests that measure mathematics reasoning on a number of achievement tests appear to directly measure quantitative reasoning (Flanagan et al., 2002). In fact, the CogAT Nonverbal score matched the predicted pattern of correlations for CogAT Quantitative score, with grades in science showing the highest correlation, followed by math, social studies, reading and writing. Further, the correlations between the CogAT Nonverbal score and grades in math and science were stronger, when compared to the correlations between the CogAT Quantitative score and grades in math and science. A potential explanation for this finding may be related to the structure of the CogAT. The CogAT Nonverbal battery purports to measure inductive reasoning skills using figural tasks, such as figures, designs, or geometric shapes as stimuli. None of the subtests on this battery require reading or the use of English. None require prior formal conceptual or factual knowledge, since all of the information needed to answer an item correctly is embedded in the item itself. By definition, the CogAT Nonverbal battery appears to be a direct measure of fluid reasoning skills or fluid intelligence. Fluid intelligence refers to mental operations that an individual uses when faced with relatively novel tasks that cannot be performed automatically. These mental operations may include forming and recognizing concepts, perceiving relationships among patterns, drawing inferences, comprehending implications, problem solving, extrapolating, and reorganizing or transforming information. Fluid intelligence also includes the narrow cognitive abilities of inductive and deductive reasoning. Based on what is reported in the literature regarding the strong relationship between fluid reasoning and mathematics achievement (e.g., Floyd, Evans & McGrew, 2003; McGrew & Flanagan, 1998), the correlation between the CogAT Nonverbal battery and grades in math and science is logical.

It is important to understand the difference between the G*f* RQ and G*q*. G*q* represents an individual's store of acquired mathematical knowledge, while RQ represents the ability to reason inductively and deductively when solving quantitative problems. G*q* would be evident when a task requires mathematical skills and general mathematical knowledge (e.g., knowing what the square root symbol means). RQ would be required to solve for a missing number in a number series task (e.g., 2 4 6 __). In sum, the CogAT Nonverbal battery appears to be a better measure of the Quantitative Reasoning (RQ) ability that is subsumed by G*f* rather than the broad ability of Quantitative Knowledge (G*q*).

(2) *Academic Achievement and Grades*

Pearson correlations found that significant relationships existed between NJ ASK scores in grade 3 and grade 7 grades in all academic subject areas investigated. The NJ ASK Language Arts score correlated the highest with grades in reading, followed by writing, social studies, science and math. This is not surprising, considering that achievement measures are related to crystallized abilities (McGrew & Flanagan, 1998), acquired knowledge (Kamphaus, 1993) and the narrower abilities associated with listening, speaking, and vocabulary development, abilities inseparable from previous experience and education. The NJ ASK Mathematics test correlated the highest with grades in science, followed by social studies, math, reading, and writing. The correlations between the NJ ASK Mathematics score and grades in science and math are logical, since both have strong foundations in mathematic concepts and formulas. While the correlation between the NJ ASK Mathematics score and grades in social studies, reading and writing is less clear, all are measures of academic performance which rely on similar mental operations, including inductive and deductive reasoning as well as forming and recognizing concepts, perceiving relationships among patterns, drawing inferences, comprehending implications, problem solving, extrapolating, and reorganizing or transforming information. Further, all demand skills in the areas of reading comprehension, written output, oral language and listening comprehension.

These findings have important implications. While research concludes that achievement test results are generally good predictors of grade point averages in high school and college samples (i.e., Kuncel, Hezlett & Ones, 2001, 2004), few research studies have looked at the relationship between standardized achievement and actual grades earned in specific subject areas. The results of this study indicate that achievement test results in grade 3 are good predictors of grade 7 grades in reading, writing, math, science and social studies, providing empirical support for this seemingly obvious relationship.

(3) *Cognitive Ability, Academic Achievement, and Predicted Achievement*

As predicted, the NJ ASK Language Arts score was the best predictor of achievement in reading and writing, when compared to the CogAT Verbal, CogAT Quantitative, CogAT Nonverbal and NJ ASK Mathematics scores. The NJ ASK Language Arts section measures students' achievement in reading and writing. Students read passages selected from published books, newspapers, magazine and everyday test and response to related multiple-choice questions, open-ended questions and a writing task. It is logical that standardized measures of reading and writing best predict actual achievement of reading and writing (i.e., grades). Further, the NJ ASK Mathematics score was the best predictor of achievement in math and science. Since the NJ ASK Mathematics section measures students' ability to solve problems by applying mathematical concepts, it is logical that that standardized measures of mathematics best predicts actual achievement in math and science. The best predictor of achievement in social studies was the NJ ASK Mathematics score, which is unexpected. In order to understand this finding, the author investigated the possibility of gender differences on the NJ ASK Mathematics score. However, the mean scale scores were not significantly different (female mean score = 232.8; male mean score = 231.7). While it is unclear as to why the NJ ASK Mathematics score best predicts achievement in social studies, the correlation (and explanation) is similar to the significant relationship found between the

NJ ASK Mathematics score and social studies grades addressed in the second question of this study. At the very least, both are measures of academic performance which rely on similar mental operations, including forming and recognizing concepts, perceiving relationships among patterns, drawing inferences, comprehending implications, problem solving, extrapolating, and reorganizing or transforming information. Further, both demand skills in the areas of reading comprehension, written output, oral language and listening comprehension.

It is interesting to note that achievement scores, rather than cognitive ability scores, were the best predictors of actual grades received. Again, this underscores the important point that above average cognitive ability may not always correlate with higher grades in specific academic subjects, as many other factors appear to play an important role in determining grades rather than ability alone.

(4) *Cognitive Ability, Academic Achievement and Special Education Classification*

Results showed that third grade NJ ASK Language Arts, NJ ASK Mathematics and CogAT Verbal scores were the best predictors of special education classification in grade 7, accounting for a combined 22% of the variance. This finding is not surprising, as students with a special education classification often present with academic deficits in the areas of reading and mathematics due to the impact of various disabilities necessary for classification (i.e., specific learning disability, other health impairment, and cognitive impairment). When children experience persistent learning difficulties, a learning disability is often suspected and a referral for a special education evaluation is frequently made. As such, scores on academic measures of reading, writing and math (i.e., NJ ASK

Language Arts and Mathematics scores) tend to be lower among the special education population in general.

What other factors account for the remaining 78% of the variance in special education prediction? It can be hypothesized that many personal characteristics other than cognitive and academic ability, including social, emotional and behavioral functioning play an important role in determining success in school. Such factors may include persistence, motivation, self-esteem, interest in school, work habits, organizational skills, study skills, interpersonal relations, attentional levels, activity levels, adaptability, anxiety, or depression. The encouragement for achievement that is received from peers, family and teachers may also be important, together with more general cultural factors.

In regards to sensitivity and specificity, the probability of correctly identifying a student as special education is low, while the probability of correctly identifying a student as not special education is high. This is not surprising, since the non-special education group is much larger than the special education group, making prediction easier and more accurate. For example, while 368 of the 379 non-classified students (97.1%) were correctly predicted as such, only 33% (N = 24) of the 73 students actually classified as special education in seventh grade were correctly predicted. In order to increase the prediction accuracy of special education classification, screenings of social, emotional, behavioral and adaptive functioning should be added to previously existing cognitive and academic measures collected in third grade. Once problem areas are identified, interventions targeting problem areas should be implemented, ideally through a tiered system.

Limitations

The most significant limitation to this study is the demographics of the sample. The participants in this study consisted of a subsample (N = 452) of the 604 seventh grade students enrolled in a large, suburban public school district in the northeastern United States. The sample was composed of a high percentage of white (89%) males (51%) and females (49%) from non-disadvantaged backgrounds (only 3.5% qualified for free/reduced lunch). Since participants were selected based on the availability of test scores from two group-administered tests over a period of five years, only those participants for whom all six scores were available were included in the analysis (N = 452). Therefore, 25% (N = 152) of the population was not included. Of the 152 participants not included in the study, demographics statistics were: 53% female, 47% male; 67% White, 11% Black, 11% unknown, 8% Asian, 3% Hispanic; 82% Not economically disadvantaged, 10% economically disadvantaged, 8% unknown; 70% not classified as special education, 29% with a special education classification. A visual comparison made between included and excluded students indicated that the excluded population has a higher percentage of students that are (1) minority, (2) economically disadvantaged, and (3) classified as special education. It is hypothesized that a majority of these students had either moved into or out of the district over the five year period. A smaller percentage of the excluded students were either absent on the day of the test and/or retest or did not take the test for other reasons, such as exemption due to severity of disability, which is related to special education classification. If these students had had a full set of scores, the results may have been different and thus more generalizable due to a more diverse population in regards to ethnicity, economic status, and special education classification.

The district's mean scores on the NJ ASK in grades 3, 5 and 7 as well as the percentage of students in the proficient and advanced proficient ranges were higher than the state means. For example, to compare the entire state of New Jersey to this district (in parentheses), on the NJ ASK 3 Language Arts test, 20.7% (7.5%) scored in the Partially Proficient range; 75.6% (87.4%) scored in Proficient range and 3.8% (5.1%) scored in Advanced Proficient range. The mean scale score in the Language Arts Literacy content area was 215.5 (225.91). For the NJ ASK 3 Mathematics test, 23.4% (10.8%) scored in Partially Proficient range; 53.8% (60.6%) scored in Proficient range and 22.8% (28.5%) scored in Advanced Proficient range. The mean scale score in the Mathematics content area was 222.2 (232.26). While the pattern of correlations may be generalizable, the data limit the degree to which the statistics can be applied to all school districts within the state. See Appendix A for a comparison of state versus district scores in grades 5 and 7.

Additionally, the students in this sample were somewhat above average in ability on the CogAT, with CogAT Composite scores ranging from 109.94 – 113.09 across the three grades. In grades 3 and 7, 96% of the students scored in the Average to Very High ranges on the CogAT Composite score. In grade 5, 97% of the students scored in the Average to Very High ranges on the CogAT Composite score. Again, these data limit the degree to which the results can be generalized beyond local norms.

A final limitation is that determination of special education status was made based on grade 7 classification. Indication of special education classification at each grade level would have been preferred. Because declassification from special education is rare, the probability of being classified in both third and seventh grade is high, which may have impacted the results of the analyses for the fourth research question.

Implications for Future Research

Future research could investigate the longitudinal relationship between the NJ ASK, CogAT and grades in middle school and outcome data in high school, such as high school grade point average, Scholastic Ability Test scores (SAT), scores on graduation exams, such as the High School Proficiency Assessment (HSPA) in New Jersey and graduation rates. Since public officials and educational administrators are increasingly calling for the use of tests to make high-stakes decisions, such as whether a student will move on to the next grade level or receive a diploma, further research in this area could identify students who score poorly on such tests and are in need of remedial services at a much earlier age. Additionally, any decision about a student's continued education, such as retention, tracking, or graduation, should not be based on the results of a single test, but should include other relevant and valid information. Further research in this area would provide information to assist in making decisions about a student's education by relying on multiple sources of data. Utilizing a more diverse sample, such as a larger sample of students nationally, would increase the generalizability of the results rather than limiting them to the local level, as was the case with the current study.

Another limitation of the current study was the collection of grades earned at the end of grade 7 only. Future research should look at the relationship between cognitive ability, standardized achievement and grades at each grade level. For example, the correlation between the grade 3 CogAT Verbal score and grade 7 reading grades was r =.33. When the correlation between the grade 7 CogAT Verbal score and grade 7 reading grades was conducted, r = .58. It can be hypothesized that the best predictor of grades would be cognitive and achievement test scores administered during the same year. This information would be particularly useful at the early grades, where important educational decisions, such as programming and even referral for special education, can be made earlier. Instead of waiting for students to fall so far behind that their performance qualifies them for special education, resources can be expended early, when problems are less intense, in the hope of remediating the problems prior to their escalation.

A final limitation is related to the cognitive measure used in this study. A drawback of the CogAT is that it was not designed to measure important cognitive factors shown in the literature to be correlated with academic achievement, other than fluid reasoning and crystallized intelligence. For example, under the broad cognitive ability area of auditory processing (G*a*), researchers have identified critical indicators that can be used to screen students for academic problems early, particularly in the areas of reading, such as phonemic awareness skills. The National Reading Panel (2000) states that in order to improve early reading skills, schools must incorporate the "essential components of reading instruction," defined as "explicit and systematic instruction in (a) phonemic awareness; (b) phonics; (c) vocabulary development; (d) reading fluency, including oral reading skills; and (e) reading comprehension strategies (National Reading Panel, 2000). One method of gathering more cognitive data regarding auditory processing skills would be the screening of these critical basic academic skills. This could be done through a tiered system and combined with previously collected data, such as the CogAT and NJ ASK scores and grades, in order to identify those students who may need additional assessments and potentially additional services within a tiered system.

Implications for Practice

The fourth research question was a preliminary investigation into the relationship between cognitive ability as measured by the CogAT and academic achievement as measured by NJ ASK test scores and special education classification. It was the investigator's hope that, by attempting to understand this relationship, group-administered cognitive and achievement scores could be added to the multiple data sources available to intervention committees when determining appropriateness of referral for a more comprehensive evaluation in order to determine special education eligibility. While the current study suggests that group-administered tests can be used to estimate future cognitive and academic ability, as well as grades, such tests are poor predictors of special education classification in later grades**.** Ideally, one would rather identify a higher percentage of students at-risk for academic failure, resulting in a higher rate of false negatives (identifying a student as not needing special education services, when in reality, they do) compared to false positives (identifying a student as needing special education, when in reality, they do not). The reverse was true in the present study.

The results of this study support the utilization of test scores in third grade as a screener to identify those children who may be at-risk for academic failure. Screening practices for education have received considerable press with the acceptance of response-to-intervention (RTI) practices in the 2004 reauthorization of the Individuals with Disabilities Education Act. The importance of screening is that students who are at risk for academic failure can be identified early and proactively. This dramatically improves the chances for successful academic outcomes for students over the traditional wait-to-fail process.

In the present study, on the NJ ASK 3 Language Arts test, 7.5% of students fell into the partially proficient range. For the NJ ASK 3 Mathematics test, 10.8% scored in partially proficient range. One fundamental assumption underlying the No Child Left Behind Act is that all students will be proficient in basic academic skills (reading and mathematics) by the school year 2013-2014. In other words, the expectation is that all students will perform to at least a basic level of proficiency. The next step is to help schools understand, with data, why their students are not proficient and what to do about it.

For example, to achieve this level of proficiency in the area of reading, it is expected that curriculum and instruction will flex and be varied based on student need. NCLB suggests that schools are to use scientific research-based practices. If instruction efforts occur early and are systemic, a significant number of students who would be considered at risk for identification of reading disability can catch up to their grade-level peers (Torgesen, 2000). Explicit and systemic instruction is a critical feature in the prevention of reading difficulties and holds the most power for preventing and remediating reading problems when combined with research-based reading practices.

By combining early screening methods within a tiered system with explicit and systematic instruction, the expectation that all students will perform to at least a basic level of proficiency may seem more realistic. This study indicates that, along with the collection of other academic and social/emotional/behavioral indicators, group-administered cognitive and academic scores are useful information to be included in early screening methods.

REFERENCES

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (1999). *Standards for Educational and Psychological Testing.* Washington, DC: American Educational Research Association.

American Psychological Association. (n.d.). *Appropriate use of high-stakes testing in our nation's schools.* Retrieved August 25, 2008, from (http://www.apa.org/pubinfor/testing.html)

Anastasi, A. (1988). *Psychological testing.* New York: Macmillan.

Benson, N. (2007). Cattell-Horn-Carroll cognitive abilities and reading achievement. *Journal of Psychoeducational Assessment*, *26*(1), 27-41.

Binet, A., & Simon, T. (1916). Upon the necessity of establishing a scientific diagnosis of inferior states of intelligence. In A. Binet & T. Simon (Eds.), (E. S. Kite, Trans.), *The Development of Intelligence in Children* (pp. 9-36). Baltimore, MD: Williams and Wilkins Company. (Original work published 1905 *L'Annee' Psychologique* (pp. 163-191).

Brody, N. (1992). *Intelligence, second edition*. San Diego: Academic Press.

Bull, R., Johnston, R. S., & Roy, J. A. (1999). Exploring the roles of the visual-spatial sketch pad and central executive in children's arithmetical skills. Views from cognitive and developmental neuropsychology. *Developmental Neuropsychology, 15,* 421-442.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* Cambridge, England: Cambridge University Press.

Carroll, J. B. (1997). The three-stratum theory of cognitive abilities. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 122-130). New York: Guilford.

Cattell, R. B. (1941). Some theoretical issues in adult intelligence testing. *Psychological Bulletin, 38,* 592.

Cattell, R. B. (1957). *Personality and motivation structure and measurement*. New York: World Book.

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

Cattell, R. B. (1967). The theory of fluid and crystallized general intelligence checked at the 5-6 year-old level. *The British Journal of Educational Psychology, 37*(2), 209-224.

Cattell, R. B. (1971). *Abilities: Their structure, growth and action*. Boston, MA: Houghton Mifflin Company.

Cattell, R. B. (1987). *Intelligence: Its structure, growth and action.* Amsterdam: North-Holland.

Cronbach, L. J. & Snow, R. E. (1977). *Aptitudes and instructional methods.* New York: Irvington.

Cummins, D. D. (1991). Children's interpretations of arithmetic word problems. *Cognitive and Instruction, 8,* 261-289.

Deno, S. L. (2003). Developments in curriculum-based measurement. *Journal of Special Education, 37*, 184-192.

Evans, J., Floyd, R., McGrew, K., Leforgee, M. (2002). The relations between measures of Cattell-Horn-Carroll (CHC) cognitive abilities and reading achievement during childhood and adolescence. *School Psychology Review, 31*, 246-262.

Felton, R. H., & Pepper, P. P. (1995). Early identification and intervention of phonological deficits in kindergarten and early elementary children at risk for reading disability. *School Psychology Review, 24*, 405-414.

Flanagan, D. P. (2000). Wechsler-based CHC cross battery assessment and reading achievement: Strengthening the validity of interpretations drawn from Wechsler test scores. *School Psychology Quarterly, 15*(3), 295-329.

Flanagan, D. P., Andrews, T. J., & Genshaft, J. L. (1997). The functional utility of intelligence tests with special education populations. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary Intellectual Assessment* (pp. 457-483). New York, NY: Guilford Press.

Flanagan, D. P., Ortiz, S. O., Alfonso, V. C., & Mascolo, J. T. (2002). *Achievement test desk references (ATDR): A comprehensive framework for LD determination.* Boston: Allyn & Bacon.

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

Floyd, R. G., Keith, T. Z., Taub, G. E., & McGrew, K. S. (2007). Cattell-Horn-Carroll cognitive abilities and their effects on reading decoding skills: g has indirect effects, more specific abilities have direct effects. *School Psychology Quarterly, 22*(2), 200-233.

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. New

York: Basic Books.

Geary, D. C. (1994). *Children's mathematical development: Research and practical applications.* Washington, DC: American Psychological Association.

Glutting, J. J., Watkins, M. W., Konold, T. R., & McDermott, P. A. (2006). Distinctions without a difference: The utility of observed versus latent factors from the WISC-IV in estimating reading math achievement on the WIAT-II. *The Journal of Special Education, 40,* 103-114.

Gottfredson, L. S. (2002a). g: Highly general and highly practical. In R. J. Sternberg & E. L. Grigorenko (Eds.), The General Factor of Intelligence: How General Is It? (pp. 331-380). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Gottfredson, L. S. (2002b). Where and why g matters: Not a mystery. *Human Performance, 15*(1-2), 25-46.

Grades 3 and 4 New Jersey Assessment of Skills and Knowledge: Technical Report. Developed by the New Jersey Department of Education, March, 2004. Retrieved on April 15, 2007 from http://www.state.nj.us/education/assessment/es/

Grades 5, 6, and 7 New Jersey Assessment of Skills and Knowledge: Technical Report. Developed by the New Jersey Department of Education, December 2006. Retrieved on April 15, 2007 from http://www.state.nj.us/education/assessment/ms/

Gray, P. (1994). *Psychology* (2nd ed.). Boston College: Worth Publishers.

Gustafsson, J. E. (1988). Hierarchical models of individual differences in cognitive abilities. In R. J. Sternberg (Ed.). *Advances in the psychology of human intelligence* (Vol. 4, pp. 35-71). Hillsdale, NJ: Erlbaum.

Gustafsson, J. E. (1994). General intelligence. In R. J. Sternberg (Ed.), *Encyclopedia of human intelligence* (pp. 469-475). New York: Macmillan.

Gustafsson, J. E. & Balke, G. (1993). General and specific abilities as predictors of school achievement. *Multivariate Behavioral Research, 28*(4), 407-434.

Gustafsson, J. E. & Undheim, J. O. (1996). Individual differences in cognitive functions. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 186-242). New York: Macmillan Library Reference USA.

Hale, J. B. & Fiorello, C. A. (2004). *School psychology: A practitioner's handbook*. New York, NY: Guildford Press.

Hale, J. B., Fiorello, C. A., Kavanagh, J. A., Hoeppner, J. B., Gaither, R. A. (2001). WISC-III predictors of academic achievement for children with learning disabilities: Are global and factor scores comparable? *School Psychology Quarterly, 16*, 31-55.

Herrnstein, R. J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American Life.* New York, NY: Simon & Schuster.

Horn, J. L. (1968). Organization of abilities and the development of intelligence. *Psychological Review, 75*(3), 242-259.

Horn, J. L. (1985). Remodeling old models of intelligence: In B.B. Wolman (Ed.). *Handbook of intelligence: Theories, measurements, and applications.* New York: Wiley.

Horn, J. L. (1988a). *Remodeling old theories of intelligence: Gf-Gc theory*. In B. B. Wolman (Ed.), Handbook of intelligence (pp. 267-300). New York: Wiley.

Horn, J. L. (1988b). Thinking about human abilities. In J. R. Nesselroade & R. B.

Cattell (Eds.), *Handbook of multivariate psychology* (rev. ed, pp. 645-685). New York: Academic Press.

Horn, J. L. (1991). Measurement of intellectual capabilities: A review of theory. In K. S. McGrew, J. K. Werder, & R. W. Woodcock, *Woodcock-Johnson technical manual* (pp. 197-232). Chicago: Riverside.

Horn, J. L. & Cattell, R. B. (1967). Age differences in fluid and crystallized intelligence. *Acta Psychologica, 26,* 107-129.

Horn, J. L. & Noll, J. (1997). Human cognitive capabilities. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 3-16). New York: The Guilford Press.

Hunt, M. (1993). *The story of psychology*. New York: Anchor Books.

Individuals with Disabilities Education Improvement Act of 2004, Pub. L. No. 108-446, 118 Stat. 2647 (2004).

Jensen, A. R. (1980). *Bias in mental testing*. New York: Free Press.

Jensen, A. R. (1984). Test validity: g versus the specificity doctrine. *Journal of Social and Biological Structure, 7,* 93-118.

Jensen, A. R. (1985). The nature of the black-white difference on various psychometric tests: Spearman's hypothesis. *The Behavioral and Brain Sciences, 8*(2), 193-258.

Kamphaus, R. W. (1993). *Clinical assessment of children's intelligence*. Boston: Allyn and Bacon.

Kaufman, A. S. (1994). *Intelligent testing with the WISC-III.* New York: John Wiley & Sons, Inc.

Keith, T. Z. (1999). Effects of general and specific abilities on student achievement:

Similarities and differences across ethnic groups. *School Psychology Quarterly, 14*, 239-262.

Kirkpatrick, E. A. (1900). Individual tests of school children. *Psychological Review, 7*, 274-280.

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2001). A comprehensive study of the predictive validity of the Graduate Record Examinations: Implications for student selection and performance. *Psychological Bulletin, 127*, 162-181.

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2004). Academic performance, creativity, and job performance: Can one construct predict them all? *Personality and Social Psychology, 86*, 148-161.

Lemaire, P. & Seigler, R. S. (1995). Four aspects of strategic change: Contributions to children's learning multiplication. *Journal of Experimental Psychology: General, 124*, 83-97.

Lohman, D. F. (2003a). The Wechsler Intelligence Scale for Children III and the Cognitive Abilities Test (Form 6): A concurrent validity study. Unpublished manuscript. Retrieved August 25, 2008, from http://faculty.education.uiowa.edu/dlohman/

Lohman, D. F. (2003b). The Woodcock-Johnson III and the Cognitive Abilities Test (Form 6): A concurrent validity study. Unpublished manuscript. Retrieved August 25, 2008, from http://faculty.education.uiowa.edu/dlohman/

Lohman, D. F., & Hagen, E. P. (2001). *Cognitive Abilities Test (Form 6).* Itasca, IL: Riverside.

Lohman, D. F., & Hagen, E. P. (2002). *Cognitive Abilities Test (Form 6): Research*

Handbook. Itasca, IL: Riverside.

Mather, N. & Jaffe, L. E. (2002). *Woodcock-Johnson III: Reports, recommendations, and strategies*. New York, NY: John Wiley & Sons.

Mather, N. & Kaufman, N. (2006). Introduction to the special issue, part two. It's about the *What, How Well,* and the *Why, Psychology in Schools, 43*, 829-834.

Mather, N. & Wendling, B. J. (2005). Linking cognitive assessment results to academic interventions for students with learning disabilities. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests and issues* (pp. 269-294). New York, NY: Guilford Press.

Mather, N. & Woodcock, R. W. (2001). Examiner's Manual. *Woodcock-Johnson III Tests of Cognitive Abilities.* Itasca, IL: Riverside.

McDermott, P. A., Fantuzzo, J. W., & Glutting, J. J. (1990). Just say no to subtest analysis: A critique on Wechsler theory and practice. *Journal of Psychoeducational Assessment, 8*, 290-302.

McDermott, P. A. & Glutting, J. J. (1997). Informing stylistic behavior, disposition, and achievement through ability subtests – Or, more illusions of meaning? *School Psychology Review, 26,* 120-142.

McGrew, K. S. (1993). The relationship between the WJ-R Gf-Gc cognitive clusters and reading achievement across the life-span. *Journal of Psychoeducational Assessment, Monograph Series: WJ-R,* 39-53.

McGrew, K. S. (2005). The Cattell-Horn-Carroll theory of cognitive abilities. In D. P. Flanagan, J. L & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 136-181). New York, NY: The Guilford Press.

McGrew, K. S. & Flanagan, D. P. (1998). *The intelligence test desk reference (ITDR): Gf-Gc cross-battery assessment.* Boston: Allyn and Bacon.

McGrew, K. S., Flanagan, D. P., Keith, T. Z., & Vanderwood, M. (1997) Beyond g: The impact of Gf-Gc specific cognitive abilities research on the future use and interpretation of intelligence tests in the schools. *School Psychology Review, 26,* 189-210.

McGrew, K. S. & Hessler, G. L. (1995). The relationship between the WJ-R Gf-Gc cognitive clusters and mathematics achievement across the life-span. *Journal of Psychoeducational Assessment, 12,* 21-38.

McGrew, K. S. & Knopik, S. N. (1993). The relationship between the WJ-R Gf-Gc cognitive clusters and writing achievement across the life-span. *School Psychology Review, 22,* 687-695.

McLaughlin, E. (2005). Relationships among Cattell-Horn-Carroll (CHC) factors of cognitive abilities and grades in high school students with learning disabilities. Ph.D. dissertation, Temple University, United States – Pennsylvania. Retrieved February 23, 2009, from Dissertations & Theses @ Temple University database.

McNemar, Q. (1964). Lost: our intelligence? Why? *American Psychologist, 19,* 871-882.

Mertler, C. A. & Vannatta, R. A. (2005). *Advanced multivariate statistical methods: Practical application and interpretation* (3rd Ed.). Los Angeles, CA: Pyrczak Publishing.

Naglieri, J. A. (2005). The Cognitive Assessment System. In D. P.

Flanagan, J. L & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 441-460). New York, NY: The Guilford Press.

National Reading Panel. (2000). *Report of the National Reading Panel: Teaching children to read.* Bethesda, MD: Author.

Neisser, U., Boodoo, G., Bouchard, T. J., Boykin, A. W., Brody, N., Ceci, S. J., Halpern, D. F., Loehlin, J. C., Perloff, R., Sternberg, R. J., & Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist, 51,* 77-101. San Diego, CA: Sattler.

No Child Left Behind Act, Pub. L. No. 107–110. (2001).

Pedhazur, E. J. (1997). *Multiple regression in behavioral research: Explanation and predictors* (3rd ed.) (pp. 261-280). Orlando, FL: Holt, Rinehart, & Winston.

Piaget, J. (1952). *The origins of intelligence in children*. (M. Cook, Trans.). New York, NY: International Universities Press.

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

Rourke, B. P. (1993). Arithmetic disabilities, specific and otherwise: A neuropsychological perspective. *Journal of Learning Disabilities, 26,* 214-266.

Rourke, B. P. & Conway, J. A. (1997). Disabilities of arithmetic and mathematical reasoning: Perspectives from neurology and neuropsychology. *Journal of Learning Disabilities, 30,* 34-46.

Sattler, J. M. (1988). *Assessment of children's intelligence and special abilities* (2d ed.). San Diego, CA: Sattler.

Sattler, J. M. (2001). *Assessment of children: Cognitive applications* (4th ed.).

San Diego, CA: Jerome M. Sattler, Publisher, Inc.

Schmitt, N., Oswald, F. L., Kim, B. H., Imus, A., Merritt, S., Friede, Al, & Fhivpuri, S. (2007). The use of background and ability profiles to predict college student outcomes. *Journal of Applied Psychology, 92*, 165-179.

Shute, G. E. & Huertas, V. (1990). Developmental variability in frontal lobe function. *Developmental Neuropsychology, 6,* 1-11.

Spearman, C. E. (1904). “General intelligence,” objectively determined and measured. *American Journal of Psychology, 15*(1), 201-293.

Spearman, C. E. (1927). *The abilities of man*. New York: Macmillan.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21*, 360-406.

Sternberg, R. J. (1985). Human intelligence: The model is the message. *Science, 230,* 1111-1118.

Sternberg, R. J., Grigorenko, E. L., & Bundy, D. (2001). The Predictive value of IQ. *Merrill-Palmer Quarterly, 47*, 1-41.

Sternberg, R. J., & Kaufman, J. C. (1998). Human abilities. *Annual Review of Psychology, 49*, 479-502.

Strang, J. D. & Rourke, B. P. (1983). Concept-formation/non-verbal reasoning abilities in children who exhibit specific academic problems in arithmetic. *Journal of Clinical Child Psychology, 12,* 33-39.

Taylor, T. R. (1994). A review of three approaches to cognitive assessment, and a proposed integrated approach based on a unifying theoretical framework. *South African Journal of Psychology, 24* (4), 183-193.

Terman, L. M. (1920). The use of intelligence tests in grading school children. *Journal of Educational Research, 1,* 20-32.

Terman, L. M. (1921). A symposium: Intelligence and its measurement. *Journal of Educational Psychology, 12*, 127-133.

The Center for Fair and Open Testing. (n.d.). *The Dangerous Consequences of High-Stakes Standardized Testing*. Retrieved August 25, 2008, from http://www.fairtest.org/dangerous-consequences-highstakes-standardized-tes

The Center for Fair and Open Testing. (n.d.). *The Testing Explosion*. Retrieved August 25, 2008, from http://www.fairtest.org/testing-explosion-0

The Public Broadcasting System. (n.d.). Can this bill live up to its' name? Retrieved August, 2008, from http://www.pbs.org/wgbh/pages/frontline/shows/schools/nochild/bush.html)

Thorndike, R. L. & Hagen, E. P. (1987). *Cognitive Abilities Test (Form 4): Technical Manual.* Chicago, IL: Riverside Publishing Company.

Thorndike, R. L. & Hagen, E. P. (1997). *Cognitive Abilities Test (Form 5): Technical Manual.* Chicago, IL: Riverside Publishing Company.

Thorndike, R. M. (1997). The early history of intelligence testing. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 3-16). New York: Guilford.

Thorndike, R. M. & Lohman, D. F. (1990). *A century of ability testing*. Chicago: Riverside.

Thurstone, L. L. (1938). Primary mental abilities. *Psychometric Monographs* (no. 1).

Torgesen, J. (2000). Individual differences in response to early interventions in reading: The lingering problem of treatment resisters. *Learning Disabilities Research & Practice*, *15,* 55-64.

Tuddenham, R. D. (1962). The nature and measurement of intelligence. In L. J. Postman (Ed.), *Psychology in the making* (pp. 469-525). New York: Knofp.

Vanderwood, M. L., McGrew, K. S., Flanagan, D. P., & Keith, T. Z. (2002). The contribution of general and specific cognitive abilities to reading achievement. *Learning and Individual Differences*, *13,* 159-188.

Vaughn, S. & Wilson, C. (1994). Mathematics assessment for students with learning disabilities. In G. R. Lyon (Ed.), *Frames of reference for the assessment of learning disabilities: New views on measurement issues* (pp. 459-472). Baltimore, MD: Paul H. Brookes.

Vernon, P. E. (1961). *The structure of human abilities* (2nd ed.). London: Methuen.

Wagner, R. K. & Torgesen, J. K. (1987). The nature of phonological processing and its casual role in the acquisition of reading skills. *Psychological Bulletins*, *101*(2), 192-212.

Wasserman, J. D. & Tulsky, D. S. (2005). A history of intelligence assessment. In D. P. Flanagan, J. L & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 3-22). New York, NY: The Guilford Press.

Wechsler, D. (1939). *The measurement of adult intelligence*. Baltimore: Williams & Wilkins.

Wechsler, D. (1958). *The measurement and appraisal of adult intelligence* (4th ed.).

Baltimore: Williams & Wilkins.

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children – Third Edition.* San Antonio: The Psychological Corporation.

Wesman, A. G. (1968). Intelligent testing. *American Psychologist, 23*, 267-274.

Williams, P. C., McCallum, R. S., & Reed, M. T. (1996). Predictive validity of the Cattell-Horn Gf-Gc constructs in achievement. *Journal of Psychoeducational Assessment, 3,* 43-51.

Wolf, T. H. (1969). The emergence of Binet's conceptions and measurement of intelligence: A case history of the creative process. Part II. *Journal of the History of the Behavioral Sciences, 5,* 207-237.

Woodcock, R. W. (1990). Theoretical foundations of the WJ-R measure of cognitive ability. *Journal of Psychoeducational Assessment, 8,* 231-258.

Woodcock, R. W. (1993). An information processing view of the Gf-Gc theory. *Journal of Psychoeducational Assessment* [Monograph Series: Advances in Psychoeducational Assessment; Woodcock-Johnson Psycho-Educational Battery – Revised], 80-102. Cordova, TN: Psychoeducational Corporation.

Woodcock, R. W. (1994). Measures of fluid and crystallized intelligence. In R. J. Sternberg (Ed.), *The encyclopedia of intelligence* (pp. 452-456). New York: Macmillan.

Woodcock, R. W. (1998). Extending Gf-Gc theory into practice. In J. J. McArdle & R. W. Woodcock (Eds.), *Human cognitive abilities in theory and practice* (pp. 137-156). Mahwah, NJ: Lawrence Erlbaum.

Woodcock, R. W. & Johnson, M. B. (1989). *Woodcock-Johnson Psycho-Educational*

Battery – Revised. Chicago: Riverside Publishing.

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

Yavanoff, P., Duesbery, L., Alonzo, J., & Tindal, G. (2005). Grade level invariance of a theoretical causal structure predicting reading comprehension and vocabulary and oral reading fluency. *Educational Measurement: Issues and Practice, 24*, 4-12.

Yerkes, R.M., Bridges, J.W., & Hardwick, R. S. (1915). *A point scale for measuring ability*. Baltimore: Warwick & York.

APPENDIX A

ADDITIONAL RESULTS TABLES

Table A-1

Percentage of Students in Each Proficient Category by Language Arts Content Area and Grade Level – State and Target School District (District)

	LAL3		LAL5		LAL7	
	State	District	State	District	State	District
AP	3.8	5.1	9.3	11.5	14.7	11.7
P	75.6	87.4	76.6	83.8	55.7	65.5
PP	20.7	7.5	14.1	4.6	29.6	22.8
Mean	215.5	225.9	220.1	226.1	215.4	218.4

Table A-2

Percentage of Students in Each Proficient Category by Mathematics Area and Grade Level – State and Target School District (District)

	Math3		Math5		Math7	
	State	District	State	District	State	District
AP	22.8	28.5	27.5	31.9	19.8	20.8
P	53.8	60.6	54.2	59.3	44.5	51.3
PP	23.4	10.8	18.3	8.8	35.7	27.9
Mean	222.2	232.3	227.6	236.0	213.3	219.7