VALIDITY OF EXECUTIVE FUNCTIONING TASKS ACROSS THE WJ III COG, NEPSY, AND D-KEFS IN A CLINICAL POPULATION OF CHILDREN: APPLICABILITY TO THREE NEUROCOGNITIVE THEORIES

A DISSERTATION

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

IN THE GRADUATE SCHOOL OF THE

TEXAS WOMAN'S UNIVERSITY

DEPARTMENT OF PSYCHOLOGY AND PHILOSOPHY

COLLEGE OF ARTS AND SCIENCES

BY

ERIN K. AVIRETT, B.A.

DENTON, TEXAS

MAY 2011

TEXAS WOMAN'S UNIVERSITY
DENTON, TEXAS

April 6th, 2011

To the Dean of the Graduate School:

I am submitting herewith a dissertation written by Erin K. Avirett entitled "Validity of Executive Functioning Tasks across the WJ III COG, NEPSY, and D-KEFS in a Clinical Population of Children: Applicability to Three Neurocognitive Theories." I have examined this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy with a major in School Psychology.

Denise E. Maricle, Ph.D., Major Professor

We have read this dissertation and recommend its acceptance:

Kimberly Booker, Ph.D.

Jenelle Fitch, Ph.D.

Daniel C. Miller, Ph.D., ABPP

Department Chair

Accepted:

Dean of the Graduate School

ACKNOWLEDGMENTS

It is first of all important to mention my appreciation of my family's, faculty's, and friend's, enduring support throughout this endeavor. To my husband, Nathaniel, your faith, wisdom, and steadfastness have been my greatest inspiration during this process. Your unfailing belief in me, as well your devotion to continuously sustain our little family have been my light. To my parents, I know that your unwavering support and constant prayers have been my source of strength during these last few years and beyond. As for my dissertation committee, Dr. Maricle, your support has extended far beyond dissertation, and even graduate school. You have never failed to motivate in times of apathy, soothe in moments of chaos, and cheerlead at every step of the way. Dr. Miller, your passion for this field and support to me, as well as all of your students, is a true inspiration. Dr. Booker and Dr. Fitch, your feedback and guidance have also been a crucial piece to this process. And finally, to my research counterpart and best friend, Jordana, my appreciation for you is beyond words. You are the only other person who will ever truly understand this experience. Thank you to all of you, I am extremely grateful for your support.

ABSTRACT

ERIN K. AVIRETT

VALIDITY OF EXECUTIVE FUNCTIONING TASKS ACROSS THE WJ III COG, NEPSY, AND D-KEFS IN A CLINICAL POPULATION OF CHILDREN: APPLICABILITY TO THREE NEUROCOGNITIVE THEORIES

MAY 2011

Inconclusive research regarding the neurocognitive construct of executive functioning has restricted the development of valid pediatric executive functioning assessments (Floyd et al., 2006; Maricle, Johnson, & Avirett, 2010). Misunderstandings in the research have led to divergent executive functioning theories and assessment tasks. Therefore, it cannot be assumed that all executive functioning instruments are measuring the same construct. Given the common inclusion of executive functioning tasks in pediatric neuropsychological evaluations (Stuss & Alexander, 2000), it is important to determine the validity of executive functioning theories and assessment tools. Furthermore, because these evaluations are often administered to children with clinical diagnoses, it is important to assess validity issues with this group. Therefore, this study aimed to determine the concurrent validity of the executive functioning subscales of three commonly utilized neurocognitive instruments: the *Woodcock Johnson III Tests of Cognitive Abilities* (WJ III COG; Woodcock, McGrew, & Mather, 2001c), the *NEPSY* (Korkman, Kirk, & Kemp, 1998), and the *Delis Kaplan Executive Function System (*D-KEFS; Delis, Kaplan, & Kramer, 2001). An associated purpose of this study was to determine the underlying factor structure of the WJ III COG, NEPSY, and D-KEFS, and their fit with three theories of executive functioning. The three theories that were analyzed include the Anderson, Levin, and Jacob (2002) model of executive functioning, the Cattell-Horn-Carroll theory of cognitive abilities (CHC theory; McGrew, 2005), and the Conceptual Model for School Neuropsychological Assessment (SNP model; Miller, 2007, 2010). Archival data was extracted from school neuropsychology case study reports. Children from a clinical sample, aged 8 through 12, were included in the study. Bivariate correlations were conducted in order to determine relationships among executive functioning subtests. These analyses revealed that executive functioning subtests appear to be measuring distinct abilities and are not interchangeable. Furthermore, the reliable use of most of these subtests within a clinical population was indicated. Level of fit between executive functioning models and sample data was depicted using structural equation modeling and analyzed using confirmatory factor analysis. The SNP conceptual model indicated the best fit with sample data.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Historically, the field of neuropsychology has focused primarily towards an adult population, which has left vast gaps in child-oriented neuropsychological research. In particular, most early neuropsychological assessments were designed exclusively for an adult population. Although a few adult instruments have been modified to also include children, they have often failed to account for the unique developmental milestones that occur during childhood and adolescence. This is an important consideration, as developmental differences among children can be substantial due to significant variability in the timing of physical, cognitive, and behavioral development during the first two decades of life.

The fields of pediatric, child clinical, and school neuropsychological assessment have recently recognized these developmental issues, as well as the idea that children are not simply small adults, but rather have unique developmental and cognitive needs (Fletcher & Taylor, 1984). Unfortunately, inconclusive research regarding certain neurocognitive abilities in children has restricted the development of valid and appropriate pediatric neuropsychological assessments. More specifically, the construct of executive functioning, or the array of cognitive abilities that make the human brain a unique, organized, and sophisticated structure, is among the least understood of the neurocognitive domains (Avirett & Maricle, 2011b). Despite limited information regarding this construct, the executive system is thought to play a critical role in human functioning, and is therefore commonly included in neuropsychological evaluations (Stuss & Alexander, 2000).

Although much empirical literature on executive functioning is available, research remains inconclusive due to inconsistent and varied findings on the topic. This lack of consensus has led to a vast misunderstanding of the structure and roles of the executive system. First of all, the physiological localization and circuitry of executive functioning skills is debated in the literature. Although executive functioning is often associated solely with the prefrontal cortex of the brain, recent research has demonstrated that numerous other structures and circuits may be involved with the executive system. For example, functional Magnetic Resonance Imaging (fMRI) research has indicated that the pre-motor cortex, as well as sections of the parietal cortex, play a role in executive functioning (Sylvester et al., 2003).

In addition to the localization debate, multiple, often competing, definitions and conceptualizations of executive functioning have been proposed in the literature (Alvarez & Emory, 2006; Anderson, Levin, & Jacobs, 2002; Baron, 2004; Boone, Ponton, Gorsuch, González, & Miller, 1998; Brocki & Bohlin, 2004; Gioia, Isquith, & Guy, 2001; Lezak, Howieson, & Loring, 2004; Stuss & Alexander, 2000; Stuss & Benson, 1986; Welsh, Pennington, & Groisser, 1991). For example, some definitions simply refer to the functioning of one behavior, such as achieving goals (Lezak et al.), whereas other definitions describe the manifestation of multiple cognitive skills. Some of the cognitive skills described include: self-monitoring and regulation, initiating and completing novel tasks, setting and achieving goals, planning, organizing, utilizing working memory, attentional control, cognitive flexibility, inhibition, and controlling emotions and social behaviors (Anderson et al.; Beaver, Wright, & Delisi, 2007; Goldberg, 2002; Hughes & Graham, 2002; Maricle, Johnson, & Avirett, 2010, Stuss & Alexander). Currently, there is a lack of consensus in the literature regarding which cognitive components or skills constitute executive functioning.

Information regarding the structure and roles of the executive system has been conceptualized into multiple executive functioning theories. Some of these models arise out of comprehensive neuropsychological theories that describe the global functioning of the brain. Within these models, executive functioning may be depicted as one of several areas of cognitive functioning. An example of a comprehensive theory of neurocognitive functioning is the Cattell-Horn-Carroll theory of cognitive abilities (CHC theory; McGrew, 2005; McGrew & Woodcock, 2001), which is a hierarchical, three-tier model of cognitive functioning. CHC theory is considered to contain the greatest breadth of comprehensive and empirically supported data of any framework available for understanding human cognitive abilities (McGrew, Schrank, & Woodcock, 2007). CHC theory asserts that an overarching "*g*" factor describes overall intelligence. Subsumed by the *g* factor are nine broad cognitive ability factors that describe general areas of cognitive processing (e.g., processing speed, short-term memory, visual spatial thinking). The lowest tier contains multiple narrow abilities that describe even more specific aspects of intelligence (McGrew; McGrew & Woodcock). Although CHC theory does not describe a specific executive functioning factor, components of the executive system are integrated into broad and narrow ability factors (Flanagan, Alfonso, Ortiz, & Dynda, 2010; Kane & Engle, 2002; McGrew & Woodcock).

Due to the extensive amount of empirical support in the literature, CHC theory has served as the theoretical foundation for multiple intellectual assessments (McGrew et al., 2007), such as the *Woodcock-Johnson III: Tests of Cognitive Abilities* (WJ-III COG; Mather & Woodcock, 2001) and the *Kaufman Assessment Battery for Children, Second Edition* (KABC II; Kaufman & Kaufman, 2004). Because of this, CHC theory influenced the development of the Cross-Battery Assessment approach (XBA; Flanagan & McGrew, 1997; Flanagan, McGrew & Ortiz, 2000). The XBA approach provides a systematic and valid way for practitioners to utilize multiple cognitive assessment tools for an evaluation, instead of being limited to only one instrument. The concept of cross-battery assessment is also an essential component of neuropsychological assessments, as multiple instruments are typically needed in order to complete a comprehensive and valid neuropsychological evaluation. However, because not all neuropsychological assessments are grounded in the same theory, cross-battery techniques utilized with neuropsychological evaluations are not as standardized as those used with XBA. Therefore, recent attempts have been made to apply CHC theory across multiple neuropsychological batteries (Flanagan et al., 2010).

Another comprehensive cognitive model, The Conceptual Model for School Neuropsychological Assessment (SNP model; Miller, 2007, 2010), focuses specifically on the assessment of neuropsychological processes in children. The SNP model organizes cognitive processes in a bottom-up approach, with higher-order cognitive functions building on more basic cognitive skills. Within the SNP model, executive functions are implicated as higher-order skills and are described as controlling sensory-motor, attentional, visual-spatial, language, and memory and learning processes. According to the SNP model, components of executive functioning that are measurable by current neuropsychological assessments are as follows: Concept Generation, Inhibition, Motor Programming, Planning/ Reasoning/ Problem Solving, Set Shifting, Retrieval Fluency, Selective/ Focused Attention, Sustained Attention, the Use of Feedback in Task Performance, and Working Memory (Miller).

As opposed to the previously discussed comprehensive models, targeted neurocognitive executive functioning theories restrict breadth of coverage exclusively to the executive system. An example of a targeted model is Anderson and colleagues (2002) model of executive functioning, which operationalizes the executive system to include the independent but interrelated components of attentional control, goal setting, and cognitive flexibility. Within this model, attentional control refers to the constructs of selective attention, sustained attention, and response inhibition. Goal setting encompasses the processes for initiating, planning, problem solving, and engaging in strategic behavior. Cognitive flexibility refers to the capacities of working memory, attentional flexibility, and self-monitoring and regulation (Anderson et al.).

Many additional comprehensive and targeted models of executive functioning have been proposed and will be discussed later in the review of the literature; however, most of these models have significant limitations and are not applicable to the current study. For example, many of the models proposed in the literature have limited research support and only include partial descriptions of the executive system. Furthermore, many of these theories fail to take the development of executive functioning into account, which is an important consideration that has only recently received significant attention in the literature.

Regarding development, it was historically thought that executive functions only emerged in late adolescence to adulthood, and therefore played no significant role in development during infancy and early childhood. Researchers are starting to recognize that the executive system plays a significant role in normal and abnormal childhood brain development (Anderson, 2002; Welsh et al., 1991). Specifically, executive functions appear to emerge in three distinct growth periods; birth to 2 years, 6 to 9 years, and adolescence to the early 20s (Anderson, 1998; Anderson et al., 2002; Hudspeth & Pribram, 1990: Romine & Reynolds, 2005).

During these growth periods executive skills develop rapidly and are particularly vulnerable to disruption (Ewing-Cobbs, Prasad, Landry, Kramer, & DeLeon, 2004). The impairment of executive skills (i.e., executive dysfunction), is evidenced by poor attention, distractibility, reduced impulse control, difficulties with planning or organization, a reduction in goal oriented behaviors, reduced insight, a tendency to blame others, or perseveration on thoughts or actions (Anderson, 2002; Bradshaw, 2001; Busch, McBride, Curtiss, & Vanderploeg, 2005). Furthermore, problems with executive functioning development may impact several childhood disorders, such as Attention Deficit Hyperactivity Disorder, Learning Disabilities, Autism Spectrum Disorders, traumatic or acquired brain injuries, epilepsy, or brain tumors (Anderson; Avirett &

Maricle, 2011a; Barkley, 2000 Brookshire, Levin, Song, & Zhang, 2004; Meltzer & Krishnan, 2007; Ozonoff & Schetter, 2007; Parrish et al., 2007; Shallice et al., 2002; Vaquero, Gómez, Quintero, González-Rosa, & Márquez, 2008).

A lack of understanding of these significant developmental considerations has led to problems regarding the assessment of these skills. Traditionally, the use of neuropsychological testing to identify the presence, absence, or structure of executive functions was focused primarily towards an adult population. This was because of the myth that executive functioning did not begin to emerge until late adolescence or early adulthood (Goldberg, 2002). Thus, very few executive functioning measures have been designed specifically for children. In fact, most assessments tools used with children are simply modifications of adult scales (Anderson, 2002; Brocki & Bohlin, 2004; Maricle et al., 2010). This is problematic, as research indicates that children and adults utilize different strategies and cognitive abilities to solve executive functioning tasks (Floyd et al., 2006).

Furthermore, extensive validity and measurement problems with many executive functioning assessments are apparent. First of all, the lack of an universal definition of executive functioning leads to an absence of any measurable, prototypical executive functioning task (Hughes & Graham, 2002), thereby making test development difficult. Additionally, most components of executive functioning involve complex tasks representing several overlapping processes indicating that any measurement of interrelated executive functioning components may be vulnerable to task impurity.

Despite these test development difficulties, several types of assessments have been proposed to identify various tasks of executive functioning. Some of these assessments may not be neuropsychological tests at all, but rather cognitive tests that have components of executive functioning built into a few subtests, such as the WJ III COG (Mather & Woodcock, 2001). Other tests are designed to assess a variety of neuropsychological tasks, and have specific executive functioning subtests that contribute to a related composite score, such as *The NEPSY, A Developmental Neuropsychological Assessment* (NEPSY; Korkman, Kirk, & Kemp, 1998). Other assessments, such as the Delis-Kaplan Executive Function System (D-KEFS; Delis, Kaplan, & Kramer, 2001), are comprised of several stand alone measures, aimed at measuring different aspects of executive functioning.

The WJ-III COG, the NEPSY, and the D-KEFS have all been utilized in the clinical setting to help identify the presence or absence of various executive functioning skills (Carper, 2003; Delis et al., 2001; Floyd et al., 2006; Korkman et al., 1998; Mather & Woodcock, 2001). However, there is a vast gap in the research regarding the validity of these measures' ability to detect executive functions. Without the substantiation that these measures are adequate identifiers of executive functioning, it is precarious to use them for this purpose in practice. In order to adequately identify executive functioning in children, the validity of these instruments needs to be assessed. Determining the validity of these assessment tools may also aid in the process of identifying the developmental pattern of executive functions' emergence in childhood.

Purpose, Rationale, and Significance of the Current Study

The purpose of the current study was to first examine the concurrent validity of subtests measuring executive functioning from three neurocognitive measurements, the WJ-III COG, the NEPSY, and the D-KEFS. This will aid in providing information regarding convergent and divergent validity of executive functioning subtests among these three batteries in a clinical population of children. Verification or contradiction regarding the validity of the WJ III COG, the NEPSY, and the D-KEFS will give practitioners information regarding which executive functioning measures for children are valid.

Specifically determining the validity of these assessments with children will aid in learning more about the nature of executive function development in children. One related and beneficial aspect to studying executive functioning in children is the potential knowledge gained about the unidentified developmental pattern of frontal lobe functions. This knowledge may also help to distinguish the specificities of various childhood behavioral and cognitive disorders (Knight & Stuss, 2002). Utilizing pure and valid measures of executive functioning, developed specifically for children, will aid in further discovering the significance of executive dysfunction in children. Furthermore, continued research and the use of more valid assessment measures will ensure more accurate clinical diagnoses, educational placements, and intervention strategies for children with executive functioning difficulties.

Additionally, the current study sought to examine the fit of various theories of executive functioning with the three assessment batteries in order to determine the efficacy of using these theories to compose assessment tools. Specifically, Anderson and colleagues' (2002) model of executive functioning, the CHC model of cognitive abilities (McGrew, 2005), and the SNP conceptual model (Miller, 2007, 2010) were analyzed. Anderson and colleagues' model was selected because it focuses specifically on executive functioning, while also encompassing a wide range of executive functioning subcomponents. The CHC model was analyzed due to its support in the literature as well as its influence on cross-battery assessment, which is an important component of neuropsychological evaluations. Finally, the SNP conceptual model was selected because of its specific focus on neuropsychological assessment in children. Moreover, a simple, one-factor executive functioning model was utilized as a comparison tool in contrast to the more advanced models.

In order to determine the fit of these models with executive functioning subtests of the WJ III COG, NEPSY, and D-KEFS, a series confirmatory factor analyses (CFA) were conducted. CFA is a statistical procedure that requires the researcher to determine the expected factor structure for a particular set of data. CFA is employed whenever the researcher has a preexisting knowledge of the underlying latent variable structure that is based on theory, empirical research, or both theory and research (Thompson, 2004). CFA is subsumed within a larger set of statistical techniques, known as structural equation modeling (SEM; Schumacker & Lomax, 2004). SEM utilizes theoretical models in order to depict relationships among observed and latent variables and establish the degree to which theoretical models are supported by sample data. Within CFA and SEM, the researcher first designates the relationships between factors and subtests in each model, based on theory. The data is then analyzed to determine how well each model or factor structure fits with sample data. In the current study, a series of CFAs were executed in order to determine the underlying structure of executive functioning tasks across the WJ III COG, NEPSY, and D-KEFS. Therefore, the following research question was addressed:

1. Is the underlying factor structure of executive functioning tasks across the WJ III COG, NEPSY, and D-KEFS in a clinical population of children best described by:
   a. A theoretical model in which all tasks load on one, general executive functioning factor?
   b. Anderson, Levin, and Jacob's (2002) theory of executive functioning, with subtests loading on the three factors of attentional control, cognitive flexibility, and goal setting?
   c. The CHC model of Cognitive Abilities (McGrew, 2005)?
   d. The School Neuropsychological conceptual model (SNP model; Miller, 2007, 2010)?

It was hypothesized that Anderson and colleagues' (2002) model of executive functioning would best fit the underlying structure of executive functioning tasks in the sample data. Specifically, this hypothesis was conjectured because the Anderson and colleagues' model is a targeted model of executive functioning that is also broad in scope. Additionally, the subcomponents included in the Anderson and colleagues' model are widely supported as important factors of executive functioning in the literature. Therefore, it was hypothesized that in comparison with the simplified one factor solution, and the broader and more complex CHC theory and SNP model, the Anderson and colleagues' model would be the best fitting and most parsimonious model.

CHAPTER II

REVIEW OF THE LITERATURE

Investigating the validity and applicability of executive functioning theories, models, and assessment tools necessitates a preexisting knowledge of several confounding factors. First of all, general knowledge of executive functioning research and associated concepts is needed. Additionally, it is important to ascertain how executive functioning research has impacted neurocognitive test development and usage. This chapter poses an overview of research related to the executive functioning construct as well as how varying theories and models of executive functioning affect the modern field of neuropsychology. It also discusses the neuroanatomy and neurodevelopment of executive skills and how those factors directly impact child development. A review of executive function impairment, or executive dysfunction, and how it relates to childhood clinical disorders is also examined. Furthermore, a review of various assessment tools geared towards the measurement of executive functioning in children, related validity research, and research regarding utilizing these measures in clinical groups of children is discussed. Next, this chapter outlines validity issues and the importance of establishing adequate validity in assessment tools. Finally, a review of the purpose and rationale of the current study is discussed.

Overview of Executive Functioning

Executive functions constitute the array of cognitive abilities that make the human brain a unique, organized, and sophisticated structure. There is no global or overarching definition of executive functions due to the abstract nature of the term and disagreement in the literature regarding the structure of executive functioning (Stuss & Alexander, 2000). However several endeavors to define the construct of executive functioning have been attempted in the literature. One commonly quoted definition is that executive functioning consists of "those capacities that enable a person to engage successfully in independent, purposive, self-serving behavior" (Lezak et al., 2004, p. 35). Another common view is that "executive functions generally refer to 'higher-level' cognitive functions involved in the control and regulation of 'lower-level' cognitive processes." (Alvarez & Emory, 2006, p. 17).

In addition to the definitional debate, several opinions as to what neurocognitive components comprise executive functioning exist. Even though proposed components of executive functioning vary widely in the literature, several constructs are generally accepted subdomains of executive functioning. Some of the commonly agreed upon executive functioning abilities are as follows: self-monitoring and regulation, initiating and completing novel tasks, setting and achieving goals, planning, organizing, utilizing working memory, attentional control, cognitive flexibility, inhibition, and controlling emotions and social behaviors (Anderson et al., 2002; Beaver et al., 2007; Goldberg, 2002; Hughes & Graham, 2002; Maricle et al., 2010, Stuss & Alexander, 2000). Higher-order components of language and the control of fine motor skills have also been attributed to executive functions. In addition, traits of emotional inhibition and response, such as tactfulness, sensitivity, and emotional affect, have also been reported to be under the control of the executive system (Goldberg; Stuss & Alexander).

The neuroanatomy and neurophysiology of executive functioning structures are also commonly debated in the literature. Historically, localization of executive functioning structures was restricted to the prefrontal cortex of the frontal lobes. However recent research has demonstrated that executive functioning activity is not limited exclusively to the frontal lobes. Although research related to this topic is diverse, a one-to-one relationship between executive functioning and the prefrontal cortex is not mutually supported in the literature (Alvarez & Emory, 2006).

The lack of consensus related to the definition, subcomponents, and structure of executive functioning plays a major role in research related to this concept, leading to differing emphases in the literature. Differences in how executive functioning is defined directly leads to conflicting theories of the construct, and thus, how executive functioning is measured and interpreted in clinical settings. These facts are complicated by the common myth in the field of neuropsychology that all executive functioning assessments are measuring the same construct. However, this is misleading, as the lack of consensus regarding executive functioning has directly led to assessment tasks that are based on divergent theories. This is why it is important to understand differences in how executive functioning is conceptualized and how differing theories and models of executive functioning are used and interpreted in clinical settings.

Conceptualization of Executive Functioning

Multiple conceptualizations regarding the structure and role of executive functioning exist in the literature. One of the most common views of the executive system delineates executive functioning as the array of processes that control and monitor specific cognitive capacities and behaviors. In this view, executive functioning is a unitary construct that functions as an overseer of specific cognitive skills. This view delineates a hierarchical conceptualization with an executive function "supervisor" overseeing narrower cognitive tasks. This viewpoint is evidenced throughout the literature, where executive functions have often been equated to both the executive of a company and the conductor of an orchestra because of the managerial skills that they are hypothesized to utilize over constituents (Goldberg, 2002).

This hierarchical theory of executive functioning stems from Baddeley and Hitch's (1974) model of working memory, which labels three primary processes involved in working memory, namely, the central executive, the phonological loop, and the visuospatial sketchpad (see figure 2.1). Within this model, the central executive works as the "supervisor" of the two other subordinate components. A third subordinate component, the episodic buffer, was later added to the model (Baddeley & Hitch, 2000). The Baddeley and Hitch (1974) working memory model directly impacted the development of the hierarchical view of the executive system, where an executive functioning construct manages multiple complex cognitive abilities.

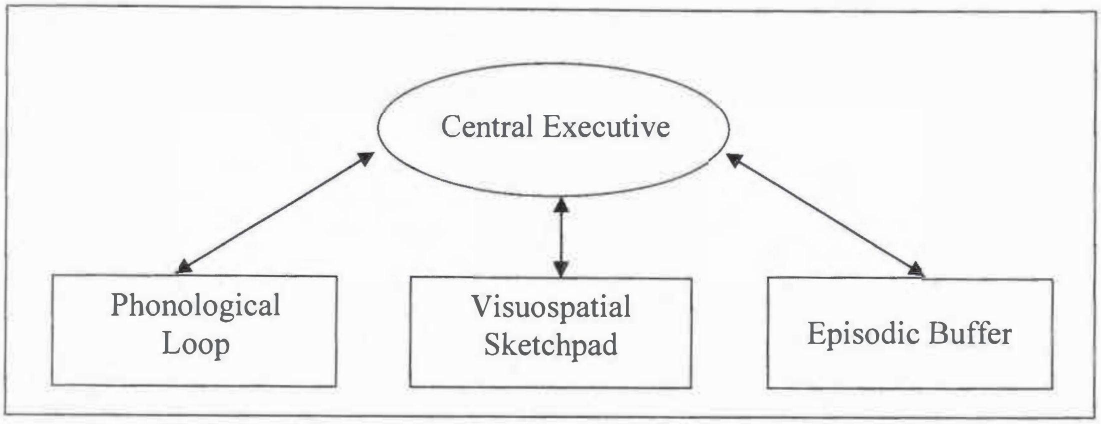

Figure 1. Illustration of Baddeley and Hitch's (2000) model of Working Memory

Within this hierarchical view, executive functioning is considered metacognitive instead of cognitive, and is often viewed as analogous to overall intelligence (Anderson, 2008; Blair, 2006; Duncan, Emslie, Williams, Johnson, & Freer, 1996; Engle, 2002; Friedman et al., 2006; Kane & Engle, 2002). Interestingly, recent research suggests that applicability of a general, one-factor conceptualization of executive functioning may be significantly affected by age and development. Using factor analytic techniques, an overarching, one-factor model of executive functioning had been found to fit well in a sample of preschool children (Wiebe et al., 2011), but did not fit to samples that included other age ranges (Latzman & Markon, 2010). One problem with the hierarchical view is that it is complex and difficult to operationally define. Therefore, directly assessing and measuring the managerial metacognitive role of executive functioning, in conjunction with the varied associated cognitive processes, is problematic. Furthermore, recent research studies have found little to no correlation between purported tasks of executive functioning and overall intelligence quotient (IQ) scores from standardized cognitive test batteries (Friedman et al.).

More recent views of executive functioning deny the hierarchical, supervisory role of the executive system, and instead state that executive functioning is simply a term for a collection of distinct and cognitively complex (i.e., higher-order) processes (Ardila, Pineda, & Rosselli, 2000; Baron, 2004; Latzman & Markon, 2010; McCabe, Roediger, McDaniel, Balota, & Hambrick, 2010). Within this view, executive functioning is purely a label for a set of associated factors and has no active, modulating role in cognitive functioning. However, even within this view, the cognitive processes that fall under the executive functioning term have yet to be decided. It is important to note that both viewpoints agree that executive functioning consists of multiple complex cognitive processes and abilities.

Theories and Models of Executive Functioning

Although there is no mutually agreed upon theory of executive functioning, several models hypothesize specific constructs and interrelated processes involved with the executive system (Busch et al. 2005). Models of executive functioning can arise out of comprehensive or targeted theories of cognitive functioning. Comprehensive neurocognitive theories broadly describe global functions of the brain, and may include descriptions of multiple neurocognitive areas, such as attention, language, memory, processing speed, and/or executive functioning. Targeted neurocognitive executive functioning theories limit breadth of coverage exclusively to the executive system.

Comprehensive neurocognitive theories. A fundamental comprehensive theory is Luria's (1980) theory of cognitive functioning, which proposes that human cognitive functions can be conceptualized within a framework of three functional but integrated units that he called “blocks.” Although Luria’s theory will not be analyzed in the current study, his theory is important to discuss as it was influential in the development of later cognitive theories and assessments, such as *The NEPSY, A Developmental Neuropsychological Assessment* (NEPSY; Korkman et al., 1998) and the *Kaufman Assessment Battery for Children, Second Edition* (*KABC-II*; Kaufman & Kaufman, 2004). Within Lurian theory, Block 1 consists primarily of the principle psychological functions that support life, such as respiration, heartbeat, and cortical arousal or attention, which are regulated by the brain stem, the diencephalon, and the medial regions of the cortex. Block II represents the areas of the occipital, parietal, and temporal lobes that are responsible for the sensory intake of information, the processing of that information, and the connections between sensory intake and other associated components of processing. Block III is involved in the regulation of the executive functions of the frontal lobes (planning, strategizing, and regulating performance) that are needed for problem solving. Block III regulates the information processed in Block II and is influenced by the basic functions of Block I (Kemp, Kirk, & Korkman, 2001; Luria).

Another comprehensive theory is the Cattell-Horn-Carroll theory of cognitive abilities (CHC theory; McGrew, 2005, McGrew & Woodcock, 2001), which characterizes an integration of two empirically supported theories of cognitive abilities. The first source stems from the psychometric factor-analytic studies of Raymond Cattell and John Horn, which became known as *Gf-Gc* theory (Horn, 1991). This theory identified several broad cognitive abilities defined as: Crystallized Intelligence (*Gc*; the breadth and depth of knowledge of a culture), Quantitative Knowledge (*Gq*; knowledge and use of quantitative facts), Reading-Writing Ability (*Grw*; reading and writing ability), Fluid Intelligence (*Gf*; novel reasoning and problem solving), Visual-Spatial Thinking (*Gv*; the ability to perceive, analyze, and think with visual patterns), Auditory-Processing (*Ga*; the ability to analyze and integrate auditory stimuli), Long-Term Retrieval (*Glr*; the ability to store information and fluently retrieve it later), Short-Term Memory (*Gsm*; the ability to hold information in immediate awareness and use it within a few seconds), and Processing Speed (*Gs*; the ability to perform automatic cognitive tasks; McGrew et al., 2007).

The second source stems from the research of Carroll (1993), which culminated in Carroll's hierarchical three-stratum theory of human cognitive abilities. Carroll identified 69 narrow abilities that he classified as Stratum I abilities. Stratum II consists of groups of narrow abilities that form several broad categories of cognitive ability, namely Fluid Intelligence, Crystallized Intelligence, General Memory and Learning, Broad Visual Perception, Broad Retrieval Ability, Broad Cognitive Efficiency, and Processing Speed. In Stratum III Carroll identified a general factor, commonly referred to as General Intelligence (*g*).

These two theories were combined to form the CHC theory of cognitive abilities (McGrew et al. 2007). CHC theory utilizes Carroll's (1993) assertion that a hierarchical, three-tier model of cognitive functioning exists and suggests that there is an overarching "*g*" factor that describes overall intelligence. Subsumed by the *g* factor are the same nine factors previously discussed within *Gc-Gf* theory namely, *Gc*, *Glr*, *Gv*, *Ga*, *Gf*, *Gs*, *Gsm*, *Gq*, and *Grw* (Horn, 1991). The lowest tier contains multiple narrow abilities that describe even more specific aspects of intelligence (McGrew, 2005; McGrew & Woodcock, 2001). This amalgamation of two theories into CHC theory is considered to contain the greatest breadth of comprehensive and empirically supported data of any framework available for understanding the organization of human cognitive abilities (McGrew et al.). Although CHC theory does not describe a specific and separate element of executive functioning, components of the executive system are integrated into the *Gf* broad ability factor (Kane & Engle, 2002; McGrew & Woodcock) and the narrow ability factors of induction, general sequential reasoning, and attention and concentration. Furthermore, Flanagan and colleagues (2010) presented multiple neurocognitive demand task analyses that outlined the loadings of various neurocognitive subtests, including measures of executive functioning, onto seven broad CHC cognitive abilities.

The CHC theory influenced the development of the Cross-Battery Assessment approach (XBA; Flanagan & McGrew, 1997; Flanagan et al., 2000), which provides a systematic and valid way for practitioners to utilize multiple cognitive assessment instruments for an evaluation. The XBA approach is also commonly utilized in neuropsychological evaluations with children (Flanagan et al., 2010). Due to its influence on the XBA approach, and thereby, neuropsychological assessment, the CHC theory was selected to be analyzed for its fit with various executive functioning tasks in the current study.

Another comprehensive cognitive model, The Conceptual Model for School Neuropsychological Assessment (SNP model; Miller, 2007, 2010), is specifically geared towards the assessment of neuropsychological processes in children. The SNP model organizes cognitive processes in a bottom-up approach, with higher-order cognitive functions building on more basic cognitive skills. Within the SNP model, sensory-motor functions and attentional processing operate as essential building blocks for all other cognitive processes and directly affect higher-order skills. The next cognitive processes, dependent on sensory-motor and attentional abilities, are visual-spatial and language processes. Memory and learning processes are the next component of the SNP model, and are dependent on the previously described domains. Executive functions are implicated next and are described as controlling sensory-motor, attentional, visual-spatial, language, and memory and learning processes. Furthermore, the speed and efficiency of processing is described as affecting all of the previously described neurocognitive functioning (Miller).

All of the described SNP domains can be subdivided into narrower abilities (Miller, 2007, 2010). Specifically, sensory-motor functions can be divided into basic sensory abilities (e.g., hearing, vision, and touch), as well as fine and gross motor skills, visual-motor integration, and balance and coordination. Attentional processes are subdivided into selective/focused attention, sustained attention, shifting attention, and attentional capacity. Visual-spatial processes are subdivided into visual perception (with and without motor response), visual-perceptual organization, and visual scanning/ tracking. Language processes are split into the following areas: phonological processing, receptive language, and expressive language. Memory and Learning is divided into the areas of immediate memory, long-term memory, and semantic memory. The speed and efficiency of processing is subdivided into the three domains of processing speed, cognitive efficiency, and cognitive fluency (Miller).

Furthermore, executive functions are subdivided into multiple components that are associated with each neurocognitive circuit of the executive system. According to Miller (2007), components of executive functioning that are measurable by current neuropsychological assessments are as follows: Concept Generation, Inhibition, Motor Programming, Planning/ Reasoning/ Problem Solving, Set Shifting, Retrieval Fluency, Selective/ Focused Attention, Sustained Attention, the Use of Feedback in Task Performance, and Working Memory. Additionally, the SNP model integrates intellectual abilities, academic abilities, and social-emotional, cultural, environmental, and situational factors into the conceptualization of a child's functioning (Miller). Due to its specific focus on neuropsychological assessment with children, the executive functioning components of the SNP model will be analyzed for fit with various executive functioning tasks in the current study.

Targeted executive functioning theories. Even with expanding knowledge regarding executive functioning, targeted models of executive functioning, which concentrate exclusively on the structure and composition of the executive system, are extremely variable in the literature. For example, Lezak et al. (2004) simply defined executive functions as the set of processes needed for engaging in independent, purposeful, self-directed behaviors. However, Gioia and colleagues (2001) and Baron (2004) identified several components of executive functioning that they organized into the following subdomains: set shifting, problem solving, abstract reasoning, planning, organization, goal setting, working memory, inhibition, mental flexibility, initiation, attentional control, and behavioral regulation.

Furthermore, Stuss and Benson (1986) illustrated a model that consists of three components: motivation, the sequencing of information, and the metacognitive control of those processes. Welsh and colleagues (1991) clustered the components of executive functions into three different factors representing speeded responding, set maintenance, and planning. Boone and colleagues (1998) specified cognitive flexibility, speeded processing, and divided attention/ short term memory as components of executive functioning. Brocki and Bohlin (2004) identified three dimensions that they interpreted as disinhibition, speed/arousal, and working memory/fluency. Busch and colleagues (2005) established three distinct factors of executive functioning that include productive fluency and cognitive flexibility, mental control and working memory, and the self-monitoring of memories.

All of the aforementioned targeted models of executive functioning describe relatively limited components of the construct, and will therefore not be included in the current study. However, in a more comprehensive model, executive functions are operationalized to include the independent but interrelated components of attentional control, goal setting, and cognitive flexibility (Anderson et al., 2002). Within this model, attentional control refers to the constructs of selective attention, sustained attention, and response inhibition. Goal setting encompasses the processes for initiating, planning, problem solving, and engaging in strategic behavior. Cognitive flexibility refers to the capacities of working memory, attentional flexibility, and self-monitoring and regulation

(Anderson et al.). Anderson and colleagues' model is a targeted executive functioning model that also represents numerous subcomponents of executive functioning; therefore, this model was selected to be included in the current study.

Methods of conceptualizing models. Several methods for conceptualizing and determining the structure of executive functioning models exist. In one method, models are based purely on theoretical views of cognitive processing (Zelazo, Müller, Frye, & Marcovitch, 2003), such as Baddeley and Hitch's (1974) model of working memory. The problem with purely theoretical models is that it is difficult to prove the validity of the theory if no quantitative data has been collected. Other models arise out of factor analytic studies that utilize neuropsychological test batteries to determine the underlying structure of executive functioning. These studies usually result in 3- or 4- factor solutions that are identified to demonstrate different aspects of executive functioning. However, the labels for each factor are subjectively decided by the researchers; therefore, the validity of these labels can be called into question (Maricle et al. 2010; Zelazo et al.).

More recent models of executive functioning have utilized brain imaging techniques in order to determine the structure of the executive system. Recently, various neuroimaging techniques, such as position emission tomography (PET), magnetic resonance imaging (MRI), and functional magnetic resonance imaging (fMRI), have been applied to the study of executive functioning. Using these techniques, suspected executive functioning subcomponents can be tested to determine if they relate to parts of the brain implicated with the executive system (Beaver et al., 2007). Most executive functioning models are either based purely on theory, or based on theory in conjunction with factor analytic studies. Only more recent studies also utilize the discussed brain imaging techniques to further support executive functioning models (Rothbart, Sheese, & Posner, 2007). Understanding how models are conceptualized and used is an important step in understanding the utility and validity of the models.

Neuroanatomy

Executive functioning processes are widely associated with the anterior most regions of the frontal lobes, known as the prefrontal cortex. However, how exactly the prefrontal cortex maintains the executive system, and the degree to which the frontal lobes regulate executive functions, is still debated in the literature (Alvarez & Emory, 2006; Hughes & Graham, 2002; Wood & Grafman, 2003). Because the frontal lobes are not coupled to any concrete or life-sustaining process, early theorists of cognitive function denied the frontal lobes any real significance, often referring to them as the "silent lobes." However, in recent years the frontal lobes have received a wide array of attention, yet remarkably, a plethora of information is still to be determined (Goldberg, 2002).

The adult human prefrontal cortex constitutes approximately one-third of the top layer of the cerebral hemisphere responsible for higher-order cognitive processes, known as the neocortex. The prefrontal cortex typically does not reach full maturation until early adulthood (Bradshaw, 2001). The prefrontal cortex accounts for 29% of the total cortex in humans, as opposed to 17% in the chimpanzee, 7% in the dog, and 3.5% in the cat (Bradshaw; Goldberg, 2002). The prefrontal cortex is considered to be the best connected region of the cortex, directly connected to every functional unit of the brain (Goldberg).

These intricate connections make the roles of executive functions exceedingly important as dysfunction in any of these areas may impact connections to other areas of the brain (Hale & Fiorello, 2004). These connections also implicate the important role that other areas of the brain play in regulating executive functions.

Executive functioning processes are split into dorsal and ventral systems working within the prefrontal cortex. The dorsal system is thought to be involved in processes related to behavioral, cognitive, and metacognitive executive function constituents. The ventral system is related to processes involved in emotional control and tone (Hale & Fiorello, 2004). Several ventral and dorsal frontal-subcortical areas originate in the prefrontal cortex and are evident during the performance of executive functioning tasks. Three commonly discussed frontal-subcortical areas include: the dorsolateral prefrontal cortex, the anterior cingulate cortex, and the orbitofrontal cortex. The dorsolateral prefrontal cortex, the last area to myelinate (i.e., insulate and speed the rate of neural connections) in the human cortex, projects to the dorsolateral head of the caudate nucleus and is associated with some of the most typical processes of executive functioning (Alvarez & Emory, 2006; Hale & Fiorello; Romine & Reynolds, 2005). Some of these processes include: cognitive and behavioral spontaneity, maintaining and shifting cognitive attention, organizational and planning strategies, goal setting, self monitoring and feedback, performing dual task activities, short-term memory, focusing and sustaining attention, response inhibition, verbal and design fluency, and regulating motor programming tasks (Alvarez & Emory; Bradshaw, 2001; Hale & Fiorello; Miller, 2007; Szameitat, Schubert, Müller, & von Cramon, 2002). The dorsolateral prefrontal cortex is also implicated in working memory, which is necessary for the temporary storage, manipulation, and retrieval of information and is theorized to be a critical component of executive functioning (Kane & Engle, 2002). Deficits within the dorsolateral prefrontal cortex are often associated with common signs of attention problems, such as inattention, poor problem solving, disorganization, and difficulties with self-monitoring and control (Hale & Fiorello).

The anterior cingulate circuit begins in the anterior cingulate, located in the medial frontal cortex area, and projects to the nucleus accumbens (Alvarez & Emory, 2006). This cortex controls behavioral processes primarily related to the initiation of and motivation for tasks and behaviors. Processes identified to be controlled by the anterior cingulate circuit include task motivation and initiation, behavioral inhibition, selective attention, working memory, self-monitoring, language, creativity, and responding to novelty (Fuster, 2002; Goldberg, 2002; Hale & Fiorello, 2004; Knight & Stuss, 2002; Miller, 2007). Damage to the anterior cingulate circuit has been associated with slow completion time for tasks, lack of persistence, apathy, limited creativity, and difficulties with self monitoring (Hale & Fiorello). Furthermore, the anterior cingulate circuit has been hypothesized to be the executive attention system, which is theorized to play a part in attentional control, working memory, and the regulation of cognitive processes and emotions (Rothbart et al., 2007). Individuals with damage to the anterior cingulate circuit often demonstrate weakness with response inhibition on neuropsychological test measures (Miller).

The orbitofrontal cortex, which originates in the ventral prefrontal cortex and projects to the ventromedial caudate nucleus, is involved with social and cognitive aspects of behavior that determine the emotional significance and social appropriateness of behavior (Alvarez & Emory, 2006; Miller, 2007). Some of the processes that are controlled by the orbitofrontal cortex include tactfulness, sensitivity, attention, emotional inhibition, planful behavior, and activity level. (Bradshaw, 2001, Knight & Stuss, 2002). Problems with the orbitofrontal cortex are associated with emotional disregulation, aggression, sexual promiscuity, disinhibition, impulsivity, and poor decision making. The orbitofrontal cortex primarily works alongside the dorsolateral prefrontal cortex, along with the anterior cingulate, in determining initiating and decision-making behavior, especially in complex problem solving situations (Hale & Fiorello, 2004).

Other less prominent connections that originate in the frontal lobes and project to other cortical, subcortical, and brainstem sites are also implicated in the executive system (Alvarez & Emory, 2006). Two such circuits are the skeletomotor circuit, which regulates large and fine muscle movements, and the oculomotor circuit, which regulates eye movements (Miller, 2007). Furthermore, another frontal circuit, the temporal/ posterior parietal circuit has been suggested to be involved with the working memory system (Miller).

Notably, the neurocircuitry of executive functioning, attention, and working memory constructs appear to overlap significantly. Specifically, research has demonstrated that the anterior cingulate circuit (Rothbart et al., 2007) and the dorsolateral prefrontal cortex (Kane & Engle, 2002; McCabe et al., 2010) are associated with all of these processes. This is a significant finding in the literature, as many researchers have historically conceptualized executive functioning, working memory, and attention as distinct systems (McCabe et al.). As was reviewed earlier, many recent theories and models of executive functioning include components of working memory and attention (Anderson et al., 2002; Baron, 2004; Boone et al., 1998; Brocki & Bohlin, 2004; Busch et al., 2005; Gioia et al., 2001). Therefore, in the current study, executive functioning, working memory, and attention will also be conceptualized as interrelated constructs.

Research further identifying the localization of constructs of executive functioning is paramount to the understanding of executive processes and related disorders. Although much has already been discovered as to the neuroanatomy and development of executive functions, there is still a lot of information yet to be uncovered. The availability of neuroanatomical studies on executive functioning is limited. Furthermore, such studies are typically narrow in focus, and only target a few subcomponents of the executive system. Although these studies are contributing to an increasing knowledge of the executive system, because of their specificity they may not generalize to the global assessment of executive skills. A more developed understanding of the neurocognitive structure of executive functioning is critical for the development of pure and valid executive functioning measures.

Neurodevelopment

Historically it has been reported that executive functions only emerge in late adolescence to adulthood, and therefore play no significant role in normal and abnormal brain development in infancy and early childhood. Only recently have researchers begun to recognize that executive dysfunction may play an important role in early childhood disorders and normal brain development (Anderson, 2002; Welsh et al., 1991). This acknowledgment has led to recent attempts to identify developmental trajectories of executive processes. Even though more research in this area is needed, studies on these developmental trajectories have been crucial in understanding executive functioning processes in children.

The development of frontal lobe activity begins to emerge as early as infancy and continues to develop throughout childhood (Welsh et al., 1991). The emergence of executive functions in childhood does not appear to be a gradual progression, but rather appears to correlate with the age dependent growth spurts of the frontal lobes. Studies have suggested that there is a significant relationship between cerebral and cognitive maturation and development (Anderson, 2002; Hudspeth & Pribram, 1990). There are major growth periods within the prefrontal cortex the first of which is birth to 2 years, followed by another growth spurt from 6 to 9 years, and a final growth period occurring from adolescence to the early 20s (Anderson, 1998; Anderson et al., 2002; Hudspeth & Pribram; Romine & Reynolds, 2005). Different factors of executive functions appear to emerge at different stages of development. This is important to take into account when working with children, as behaviors that are typically associated with executive dysfunction, such as cognitive inflexibility, disinhibition, and working memory deficits, can be either appropriate or problematic, depending on the child's developmental stage.

Birth and infancy. At birth, the primary areas of the brain responsible for daily living functions, such as breathing, eating, and regulatory processes, as well as the associated connections to the prefrontal cortex, are operational (Romine & Reynolds, 2005). At this time, the functioning of the prefrontal cortex appears to be minimal compared to other more developed brain areas; however, executive functioning pathways and connections are still present. Notably, some of these pathways may not finish myelinating until childhood, adolescence, or even early adulthood, thereby making them less efficient during infancy.

During this time period, infants are rapidly learning new information, strategies, and motor planning skills that aid in getting crucial needs fulfilled. For example, infants are able to visually search for toys or a caregiver, perform simple behavioral inhibition (i.e., refrain from touching something hot), and can learn a caregiver's non-verbal cues (Dawson & Guare, 2009). Therefore, during early infancy there is evidence for the executive functioning processes related to learning, organized search, basic inhibitory skills, goal-directed behavior, and expansive memory and attentional skills. Infants are also learning social cues and nonverbal signals that allow them to adapt and connect to caregivers (Romine & Reynolds, 2005; Welsh et al., 1991). However, it is important to note that validly measuring these skills in infants is challenging; therefore, an understanding of the full range of executive functioning abilities in infancy has yet to be determined.

Early and middle childhood. The most significant period of executive functioning development appears to occur in the early to middle childhood years. These periods of growth seem to mirror expansion of general cognitive abilities during this time. During the early to middle childhood period multiple cerebral changes occur, such as the pruning of neuronal synapses. Pruning, or the specific elimination of unneeded neuronal connections, can increase and accelerate synaptic transmission (Luna, Garver, Urban, Lazar, & Sweeney, 2004). In addition, the continuation of prefrontal myelination can also make neuronal transmission faster and more effective (Romine & Reynolds, 2005). These processes make cognitive processing more efficient and directly affect the development of associated executive functioning skills.

Multiple executive functioning processes related to the control of goal-directed behaviors are beginning to advance during this time period. For example, most children in this developmental range are able to complete multi-part directions, engage in behavioral inhibition (e.g., raising a hand before speaking), complete homework assignments independently, and perform simple chores and self-help tasks without reminders (Dawson & Guare, 2009). Therefore, some of the specific executive functioning processes that may develop during this time include: concept formation, creative problem solving, working memory, planning, organizational strategies, and inhibitory skills (Romine & Reynolds, 2005; Welsh et al., 1991). However, it is important to keep in mind that adult level maturation of these executive functioning skills might not occur until adolescence or adulthood.

Late childhood, adolescence, and early adulthood. During late childhood, adolescence, and early adulthood cortical changes occur as children continue to develop. Development of the prefrontal cortex, through additional pruning and myelination, is said to reach maturation around late adolescence; however, certain aspects of this maturation may continue into early adulthood (Goldberg, 2002; Romine & Reynolds, 2005).

Pathways within the dorsolateral prefrontal cortex are some of the last to fully myelinate in the adult human cortex (Miller, 2007; Romine & Reynolds), and several executive functioning processes related to the dorsolateral prefrontal cortex have been reported to continue developing after the age of twelve.

During this time period most adolescents are able to drive, keep track of changing daily schedules, and plan for long-term goals (Dawson & Guare, 2009). The executive functioning components that impact these behaviors include: the advanced coordination of working memory and inhibition, complex planning and foresight, visual working memory, verbal fluency, processing speed, and motor sequencing (Anderson et al., 2002; Luna et al., 2004; Romine & Reynolds, 2005). However even in adolescence not all executive functioning skills have fully matured. Increased emotional arousal, paired with immature behavioral inhibition and self-control, often leads to increased risk-taking behaviors in adolescence (Figner, Mackinlay, Wilkening, & Weber, 2009; Steinberg et al. 2008; Young et al., 2009). However, by early adulthood all components of the executive system should be fully developed and active in the typically developing individual.

Executive Dysfunction and Clinical Disorders in Children

The executive system is extremely important in daily functioning, as impairment to the regions of the brain responsible for these processes can result in severe difficulties in everyday cognitive and behavioral tasks (Bradshaw, 2001; Goldberg, 2002). Executive functions are thought to mature in rapidly developing age spurts, during which they are vulnerable to disruption (Ewing-Cobbs et al., 2004). Difficulties associated with components of the executive system are often referred to as executive dysfunction.

Executive dysfunction is an important research topic, as much of what science has learned about executive functions is through what is known of executive dysfunction (Avirett & Maricle, 2011a). Executive dysfunction may materialize from congenital abnormalities or disorders, traumatic or acquired brain injury, or brain lesions (Stuss & Alexander, 2000). Specifically, executive dysfunction can result if damage occurs in one of the frontal-subcortical areas of the anterior most regions of the brain, known as the prefrontal cortex (Bradshaw; Goldberg). Executive dysfunction can manifest by poor self control, poor planning or organizational skills, difficulty generating individual strategies for problem solving, or perseveration on thoughts or ideas (Anderson et al., 2002). Executive dysfunction may also be evidenced by poor attention and distractibility, difficulty inhibiting responses, a reduction in self-generated behaviors, reduced insight, a tendency to blame others, or difficulty learning from past experiences (Bradshaw; Busch et al., 2005).

Additionally, researchers have recently discovered that problems with executive functioning development may significantly impact several childhood disorders (Anderson, 2002). Attention Deficit Hyperactivity Disorder (ADHD), marked by problems with inhibition and the processes of attention, has been linked to executive dysfunction in children (Barkley, 2000; Shallice et al., 2002). Although deficits in specific components of executive functioning, such as response inhibition, vigilance, working memory, and planning, have been implicated in ADHD, global deficits in executive functioning are also apparent (Willcutt, Doyle, Nigg, Faraone, & Pennington, 2005).

Furthermore, executive dysfunction has been associated with children with learning disabilities due to cognitive weaknesses that directly affect academic learning. These weaknesses include problems with self-regulation and monitoring, problem solving, cognitive flexibility, and organizing and prioritizing stimuli (Meltzer & Krishnan, 2007). Executive dysfunction has also been implicated in nonverbal learning disabilities because of associated difficulties with cognitive flexibility, fluently shifting tasks and environments, adapting to novel situations, working memory, self-regulation, and attentional control (Stein & Krishnan, 2007).

In addition, executive dysfunction has been implicated in children diagnosed with autism spectrum disorders. Children with autism spectrum disorders may have problems with cognitive flexibility, fluently shifting attention to new or novel tasks, planning, appropriately responding to social cues, regulating social interactions, and nonverbal behaviors (Ozonoff & Schetter, 2007). Although executive dysfunction plays a significant role in autism spectrum disorder, manifestations of executive dysfunction do not clearly define differences in autism spectrum subtypes (Verté, Geurts, Roeyers, Oosterlaan, & Sergeant, 2006).

Executive dysfunction may also be apparent in children with traumatic or acquired brain injuries, epilepsy, or brain tumors. In these children, impairments in executive functioning are often associated with age of onset, as well as the severity and location of injury, seizure activity, or tumor. However, global impairments in executive functioning, as well as acute disruptions in attention and processing speed, are common (Brookshire et al., 2004; Parrish et al., 2007; Vaquero et al., 2008).

The review of the literature on executive dysfunction and clinical disorders reveals that some differences in the clinical manifestation of executive functioning impairments exist. However, children diagnosed with associated clinical disorders (i.e., ADHD, learning disabilities, autism spectrum disorders, traumatic/ acquired brain injuries, epilepsy, brain tumors, etc.) often demonstrate global executive functioning impairments that do not distinguish clinical groups (Anderson, 2002; Ozonoff & Schetter, 2007). This is an important consideration for individuals assessing for executive dysfunction in clinical groups of children.

Assessment of Executive Functioning

Although there are many neuropsychological tests aimed to measure tasks of executive functioning, multiple assessment problems regarding these tasks in children have ensued. First of all, the development of executive functioning measures has been primarily focused towards an adult population. This may be because early neuropsychological assessments designed for adults were complex and difficult to modify or administer to children. Additionally, the incomplete development of executive functions and limited understanding of the manifestation of these abilities in children has made developing executive functioning tasks for a younger population difficult. Therefore, very few measures of executive functioning that have been designed specifically for use with children exist. Most of the neuropsychological test instruments that are currently available for use with children are modifications of similar adult tasks and lack documented indices of reliability and validity with children (Delis et al., 2001; Humes, Welsh, Retzlaff, & Cook, 1997; Maricle et al., 2010). This is problematic, as research indicates that children and adults utilize different strategies and cognitive abilities to solve executive functioning tasks (Floyd et al., 2006). Furthermore, most research regarding the validity of executive functioning assessments only involves adult-level tasks, leaving a void in the availability of research-validated executive functioning tasks for children.

Furthermore, general validity and measurement problems with many executive functioning assessments are apparent. One problem associated with executive functioning assessments is in determining which aspect of the executive system should be measured. Most components of executive functioning involve complex tasks representing several overlapping processes. This denotes that any measurement of interrelated executive functioning components may be vulnerable to task impurity. Furthermore, the lack of an operational definition of executive functioning leads to an absence of any measurable, prototypical executive functioning task (Hughes & Graham, 2002), thereby making test development difficult. Also, since the prefrontal cortex works in conjunction with many other areas of the brain, it is difficult to identify which region of the brain is responsible for outcomes on executive functioning measures (Fuster, 2002; Goldberg, 2002; Hudspeth & Pribram, 1990). All of these issues have contributed to a general trend of low test-retest reliability and variable validity in executive functioning measures.

In addition to the measurement issues already discussed, many myths regarding the assessment of executive functioning exist within the field (Maricle et al., 2010). One early myth was that impaired performance on tests of executive functioning directly implicated dysfunction of the frontal lobes (Bradshaw, 2001; Goldberg, 2002). Another early myth related to executive functioning assessment was that each assessment task corresponded to a single executive function subcomponent (Anderson, 2002). Recent research has disproven both of these myths by demonstrating that adequate performance on tasks of executive functioning require multiple cognitive skills such as attention, perception, concept formation, working memory, inhibition, planning, and cognitive flexibility (Ardila et al., 2000; Baron, 2004; Latzman & Markon, 2010; Maricle et al.; Roediger et al., 2010). These cognitive skills are not just restrained to the frontal lobes, but rather identified in multiple areas and structures of the brain (Fuster, 2002). Deficient performance on tasks of executive functioning can result at multiple stages of cognitive processing, from basic cognitive skills (e.g., attention, perception), to higher-level functions (e.g., cognitive flexibility, working memory) (Ardila et al.; Delis et al., 2001).

Although the measurement of executive functioning tasks has proven to be difficult, several notable neuropsychological tests designed to evaluate these skills have been developed. Even though no instrument has been developed to assess the entirety of the executive system, many assessment tools intended to measure specific aspects of executive functioning are available. For many executive functioning measures, issues of reliability and validity still need to be researched.

Foundational Executive Functioning Assessments

Several foundational, stand-alone measures of executive functioning will be briefly discussed. Although many of these tasks are still utilized in clinical practice, they have served as a foundational blueprint for current re-normed and updated assessment batteries, such as the *Delis-Kaplan Executive Function System* (D-KEFS; Delis et al., 2001). These measures have a lengthy history, some of which were developed and put into practice as early as the 1930s. Due to the historical significance of these tasks, many of these assessments already have a large research basis. Some of the most widely researched tasks include the *Stroop Color-Word Test* (Stroop, 1935), the *Wisconsin Card Sort Test Revised and Expanded* (WCST; Heaton, Chelune, Talley, Kay & Curtiss, 1993), and various category, trail-making, and tower tests. Although this list is not comprehensive, it includes some of the most researched and widely used tasks. All of the aforementioned assessments are stand-alone tasks that measure a single element of executive functioning. On account of their limited scope, these tasks will not be utilized in the current study. However, it is still important to discuss these measures due to their influence on the development of more comprehensive executive functioning tasks.

Stroop Color-Word Test (SCWT). The Stroop procedure is one of the oldest, most popular, and most prevalent techniques used for assessing higher-order cognitive skills. The *Stroop Color-Word Test* (SCWT; Stroop, 1935) was developed to determine inhibitory response skills in children and adults. The Stroop task compels the respondent to inhibit and replace verbalizing a well-learned response with a novel response (e.g., inhibit saying the name of a color by verbalizing the name of a competing color). It is thought to measure focused attention, selective attention, mental flexibility, and response inhibition. Several neuroimaging studies have implicated activation of the superior medial prefrontal cortex, anterior cingulate circuit, and cerebellum during the inhibition condition of the Stroop task (Gruber, Rogowska, Holcomb, Soraci, & Yugerlun-Todd, 2002; Pujol et al., 2001).

The *Stroop Color and Word Test-Revised, Children's Version* (Golden, Freshwater, & Golden, 2003) is an updated, norm-referenced, standardized version of the classic Stroop task. During most Stroop tasks, the participant is first administered one or two baseline conditions (e.g., name color patches or color words). Next, the examinee is administered the experimental condition in which they have to inhibit their well-learned response (e.g., reading the color of the ink, rather than the dissonant printed color word). Some Stroop tasks incorporate an additional condition in which the examinee is required to switch between naming the dissonant ink color and reading the words. Many current neuropsychological test batteries, such as the D-KEFS and the NEPSY-II include Stroop tasks.

Research on Stroop tasks has indicated that not all Stroop tasks are interchangeable and may not measure the same neurocognitive processes (Salthouse & Meinz, 1995; Shilling, Chetwynd, & Rabbitt, 2002). Although there is a significant research base for Stroop tasks, the lack of consistency between tasks makes finding trends in the literature difficult. A meta-analysis of the *Stroop Color and Word Test* with children (Homack & Riccio, 2004) indicated that most research on Stroop tasks is with adults. Not only is research on Stroop tasks in children limited, but no normative base for children exists, which makes conceptualizing the effects of Stroop tasks in children difficult. The meta-analysis revealed that the *Stroop Color and Word Test* effectively discriminates children with ADHD and Learning Disabilities from typically developing peers. However, it is not sensitive in determining differences between clinical groups of children (Homack & Riccio).

Wisconsin Card Sort Test Revised and Expanded (WCST). The *Wisconsin Card Sort Test Revised and Expanded* (WCST; Heaton et al., 1993), originally published in 1981, is one of the most widely used tests of executive functioning in adult populations. The WCST is considered to be a measure of concept formation, abstract reasoning, response inhibition, sustained attention, cognitive flexibility, and problem solving. The WCST is a sorting task that requires the participant to form, shift, and maintain concepts and conditions. The current version of the WCST was designed for implementation with individuals ages 6 to 89 years of age. During administration of the WCST respondents are asked to determine the rule among a series of cards. After a predetermined amount of trials the rule is changed without the examinee's knowledge; however, the individual is required to determine the new rule.

Research involving the WCST indicates that performance on the task steadily increases from age 6.5 through age 19, remains steady through age 50, and then begins to decline. Some research has demonstrated that adult level performance is achieved by age 10 on the WCST (Welsh et al., 1991). Research with above average and gifted children, children with traumatic brain injuries, and children with ADHD indicates that performance on the WCST is positively correlated with overall intelligence (Arffa, Lovell, Podell, & Goldberg, 1998; Slomine et al., 2002). However, these results directly challenge those of Welsh and colleagues and Levin and colleagues (1991), which found no relationship between performance and intelligence.

In a meta-analytic review of 32 journal articles and one dissertation (Romine et al., 2004), results of the WCST administered to children with specific clinical disorders (e.g., ADHD, Learning Disorder, Conduct Disorder, Autism Spectrum Disorders, Mood Disorders, and schizophrenia) were analyzed. The results suggest that individuals with Learning Disorders, Conduct Problems, and Autism Spectrum Disorders consistently performed worse than controls. Results for children with ADHD were mixed. Although students with ADHD performed worse than controls, they performed better than other clinical groups. Children with anxiety disorders demonstrated more total errors and perseverative responses than controls. Results for children with depression were mixed, as some studies identified children with depression to demonstrate significant deficits on the WCST, whereas other studies did not find such results. Several studies demonstrated that deficits were apparent in children with schizophrenia; however, due to a limited amount of studies on schizophrenia and the WCST in children, results were left inconclusive.

Category tests. Category tests are measures of executive functioning that are considered to measure concept formation, mental shifting, rule learning and problem solving (Baron, 2004). Similar to sorting tasks, category tests determine the ability to use trial and error skills to solve problems and to utilize feedback to determine the validity of hypotheses. Individuals who struggle with category tests may manifest limited cognitive flexibility, learning inefficiency, and potential memory problems. The original Category test was developed as part of the *Halstead-Reitan Neuropsychological Test battery* in the 1940s and was utilized to determine the ability to develop strategies based on past experiences and the integration of new information (Boll, 1993). Although the original Category Test was designed for use with adults, downward extensions were later developed for the *Halstead-Reitan Neuropsychological Test Battery for Older Children* and the *Reitan-Indiana Neuropsychological Test Battery for Young Children* (Reitan & Wolfson, 1992). Following in the theoretical footsteps of these tests, the *Children's Category Test* (CCT; Boll) was developed to assess concept formation, problem solving, and nonverbal learning and memory in children ages 5 though 16. Although the CCT was developed specifically for use with children, research has demonstrated that age significantly affects outcomes on the task, so an age-correction formula is needed (Donders, 1998).

Trail-making tests. Trail-making tests are commonly used in neuropsychological batteries for children and adults for measuring executive functioning (Horowitz, Schatz, & Schute, 1997). They were originally created as part of the classic *Army Individual Test Battery* by Partington and the U.S. Army (1944), and were later included in the *Halstead-Reitan Neuropsychological Battery* (Halstead, 1952). Since this time, trail-making components have been introduced in other neuropsychological assessment batteries, such as the D-KEFS. Trail-making tests are purported to measure visual attention, visual perception, inhibition, and cognitive processing speed. Furthermore, trail-making tests are implicated as valid measures of broad cognitive functioning because they require left hemispheric functioning, due to a focus on symbolic recognition, as well as right hemispheric functioning, due to visual scanning (D'Amato & Hartlage, 2008).

Most trail-making tasks are comprised of multiple components. The first section typically serves as the baseline for visual scanning and tracking by requiring the examinee to sequence items quickly. Then, the second task measures cognitive flexibility and switching on a visual-motor sequencing task. A disadvantage of trail-making tasks is that it is difficult to determine if poor performance is the result of executive dysfunction, poor processing speed, poor visual scanning, poor sequencing abilities, or poor fine motor coordination (D'Amato & Hartlage, 2008; Miller, 2007). This is accounted for in some trail-making tasks, such as in the Trail Making Test of the D-KEFS, by utilizing multiple trials that help partial out executive skills from processing speed and motor control skills. Another disadvantage is that most trail-making tasks used with children are simply downward extensions of similar adult tasks. It is not clear whether these downward extensions measure the same constructs in children as they measure in adults.

Tower tests. Tower tests typically measure planning, working memory, response inhibition, and visuospatial memory (Maricle et al., 2010). Most tower tests require the examinee to arrange objects from an initial position to a new position, with a pre-determined number of moves and rules. Multiple versions of tower tasks have been designed, including the *Tower of London* (Shallice, 1982), *Tower of London-Drexel University* (Culbertson & Zillmer, 1998), the *Tower of Hanoi* (Simon, 1975), and the Tower Test from the D-KEFS (Delis et al., 2001). In addition, the original NEPSY also contained a tower task (Korkman et al., 1998), but it was not included in the updated NEPSY-II (Korkman, Kirk, & Kemp, 2007a, 2007b). Each version of the tower task differs in the way the task is structured, the rules for completing the task, and the way the task is measured and evaluated (D'Amato & Hartlage, 2008).

Research suggests that tower tasks should not be used interchangeably due to differences in the structure and neurocognitive demands imposed by each task (Baron, 2004). Furthermore, several studies (Goel, Pullara, & Grafman, 2001; Humes et al., 1997) have revealed that the *Tower of London* and *Tower of Hanoi* tasks do not correlate and appear to be measuring different cognitive abilities. The *Tower of Hanoi* appears to measure response inhibition more than planning abilities. Further research suggests that tower tasks might not be useful when assessing executive functioning in children, because of differences in the skills used by children and adults when completing the tasks (Baker, Segalowitz, & Ferlisi, 2001; Bishop, Aamodt-Leeper, Creswell, McGurk, & Skuse, 2001).

Children's performance on tower tasks appears to be impacted by developmental differences (Culbertson & Zillmer, 1998). Overall, typically developing children (ages 7-15) demonstrate improved performance and problem-solving proficiency with age, and progressively decreased frequency of rule violations. Younger children (ages 7-9) demonstrate limited planning, frequent rule violations, and problem solving by trial-and-error methods. Older children (ages 10-12) show more inconsistent planning abilities, in that they show advanced strategies on simpler problems but immature strategies on more complex problems. Adult levels of performance are reached around age 13 to 15. Impaired performance on the *Tower of London* has been reported in the literature for several childhood disorders, including traumatic brain injury, hydrocephalus, brain lesions, and phenylketonuria (Fletcher, Brookshire, Landry, & Bohan, 1996; Jacobs & Anderson, 2002; Levin et al., 1991; Welsh, Pennington, Ozonoff, Rouse, & McCabe, 1990). While research has demonstrated that tower tasks are sensitive to brain impairment, they are not specific as to the localization of the impairment or their ability to distinguish between clinical groups of individuals.

Comprehensive Executive Functioning Tasks

As previously discussed, all of the foundational executive functioning assessments consist of stand-alone tasks that measure a select element of executive functioning. Many of the foundational executive functioning assessments have been updated, re-normed, and assembled with other executive functioning tasks to form more current and comprehensive test batteries. Different types of comprehensive assessment batteries include measures of executive functioning (Maricle et al., 2010). Some of these tasks are cognitive batteries that have elements of executive functioning built into certain subtests, such as the *Woodcock Johnson III: Tests of Cognitive Abilities, Normative Update* (WJ III-COG; Woodcock, Schrank, Mather, & McGrew, 2007). Other assessment instruments were developed as comprehensive neuropsychological batteries and contain specific executive functioning subtests that contribute to a related composite score, such as the NEPSY and NEPSY-II (Korkman et al., 1998, 2007a, 2007b). Additionally, some of these assessments are designed to contain several stand-alone measures, standardized together in one test battery, such as the *Delis-Kaplan Executive Function System* (D-KEFS: Delis et al., 2001).

Although additional neurocognitive batteries that include executive functioning components currently exist, the WJ III COG, NEPSY, and D-KEFS will be the focus of the current study due to theoretical or foundational basis of each battery. Additionally, the purpose of focusing on these three batteries is to widen the currently sparse literature on executive functioning tasks among the WJ III COG, NEPSY, and D-KEFS.

Woodcock Johnson III Tests of Cognitive Abilities, Normative Update (WJ III COG NU). The *Woodcock Johnson III Tests of Cognitive Abilities, Normative Update* (WJ III COG; Woodcock, McGrew, & Mather, 2001c; Woodcock et al., 2007) is a comprehensive set of individually administered, norm-referenced tests for measuring intellectual abilities in individuals ages 2 to 95 (Schrank, 2005). The WJ III COG and its predecessors were developed to correspond with a specific theoretical orientation of cognitive abilities that came to be known as the Cattell-Horn-Carroll (CHC) theory, which asserts that a hierarchical, three-tier model of cognitive functioning exists.

Mather and Woodcock (2001) utilized CHC theory as the blueprint for the WJ III COG. Included in the WJ III COG are 20 subtests that are each interpreted to represent a distinct and unique narrow ability. Combinations of these subtests provide cluster scores that characterize seven of the nine broad ability areas of CHC theory. The seven broad cluster abilities included in the WJ-III COG are: Comprehension-Knowledge (*Gc*), Long-Term Retrieval (*Glr*), Visual-Spatial Thinking (*Gv*), Auditory Processing (*Ga*), Fluid Reasoning (*Gf*), Processing Speed (*Gs*), and Short-Term Memory (*Gsm*). The WJ-III COG also provides and strongly emphasizes an overall general measure of *g*, represented by the General Intellectual Ability (GIA) score.

In addition, the WJ-III COG includes specific clusters representing broad factors of cognitive abilities that are related to cognitive performance, namely Acquired Knowledge, Thinking Ability, and Cognitive Efficiency. Several additional clusters may also be obtained from other combinations of subtests within the WJ III COG that are useful in the diagnostic or clinical setting, these clusters include Phonemic Awareness, Working Memory, Broad Attention, Cognitive Fluency, and Executive Processes (Mather & Woodcock, 2001). The Executive Processes and Broad Attention clusters are designated to measure aspects of executive functioning (Mather & Woodcock).

The Executive Processes cluster in the WJ III COG incorporates three aspects of executive functioning, including strategic planning, interference control (response inhibition), and mental (cognitive) flexibility. The three tests that form this cluster are Concept Formation, Planning, and Pair Cancellation. Concept Formation is a controlled learning task that requires rule formation after a specific set of stimuli has been presented. This subtest measures the controlled ability to alter one's mental set (Mather & Woodcock, 2001). There is no memory component to this subtest as there is a stimulus key presented throughout. This task also utilizes induction, categorical reasoning, and logic. Within the Planning subtest the examinee is required to employ forethought and mental restriction in order to completely trace a given stimulus without removing the pencil from the piece of paper or retracing any lines. The narrow abilities that the Planning subtest requires are spatial scanning and general sequential reasoning. The Pair Cancellation subtest is a timed test that requires the examinee to locate and mark a specific pattern of repeating objects on a provided stimulus card. It requires the capacity to stay on task in a vigilant manner and involves the narrow abilities of attention, concentration, and interference control (response inhibition).

The Broad Attention cluster assesses a global overview of attention (Mather & Woodcock, 2001). The four subtests within the Broad Attention cluster, Numbers Reversed, Auditory Working Memory, Auditory Attention, and Pair Cancellation, each measure different aspects of attention. In Numbers Reversed the subject mentally holds a series of numbers while reversing the sequence. It requires attentional capacity, working memory, and transformation. In Auditory Working Memory, the subject listens to a series of numbers and words, and then repeats the words in sequential order, followed by the numbers in numerical order. Auditory Working Memory necessitates divided attention, working memory, reorganization, sorting, and sequencing. In Auditory Attention, the subject is asked to listen to phonetically similar words (i.e., bee, knee, sea), while presented with increasingly intense background noise. Subjects are then required to point to a picture that represents the word they heard. This task entails selective attention, speech-sound discrimination, and inhibition of extraneous auditory stimuli. As already discussed with the executive processes cluster, Pair Cancellation requires sustained attention.

Although a plethora of research utilizing WJ III COG components exists, most of this research focuses on the broad CHC factors. Therefore, limited research regarding the Executive Processes and Broad Attention Clusters exists in the literature (Maricle et al., 2010). The WJ III COG has also been widely studied within populations consisting of individuals with clinical diagnoses (e.g., ADHD, Autism Spectrum Disorders, Specific Learning Disabilities). Although these studies have demonstrated diagnostic validity and utility with these populations, they have focused on the validity of the broad ability factors instead of the Executive Processes or Broad Attention Clusters.

NEPSY and NEPSY II. The NEPSY (Korkman et al., 1998) and its revision, the NEPSY-II (Korkman et al., 2007a, 2007b), were developed as neuropsychological assessment tools designed specifically for children. The NEPSY was designed for children ages 5 to 12, and the NEPSY-II extended the age range to 3 through 16. The NEPSY, as opposed to the NEPSY II, will be included in the current study due to the use of archival data, and the inclusion of a Tower task in the NEPSY. Kemp et al. (2001) emphasize that the NEPSY was developed with four associated purposes in mind. The first purpose was to create a valid and reliable neuropsychological instrument for children designed to be sensitive to skills across five functional domains. The second purpose was to contribute to the knowledge base of congenital or traumatic brain damage. The NEPSY was also intended to be used in long term follow up of children with brain damage or dysfunction. Finally, the NEPSY was intended to study neuropsychological development in preschool and school-aged children.

Development of the NEPSY was based upon the theoretical work of Luria (1980), whose research has been considered to be a major foundation for neuropsychology for nearly four decades (Kemp et al., 2001). As previously discussed, Luria theorized that human cognitive functions can be conceptualized within a framework of three functional but integrated units that he called "blocks". Each block is composed of cognitive processes and increase in complexity with progression through the blocks (Kemp et al.; Luria, 1980). Luria also proposed that impairment in one particular function will affect other complex and connected cognitive functions because all areas are interconnected or intercorrelated. Luria's approach focuses on identifying the primary deficits involved with impaired performance in one function which may then contribute to a secondary deficit in another functional domain (Kemp et al.; Korkman, 1988).

The NEPSY and NEPSY-II were inspired both by Luria's approach to assessing cognitive functions and the need for neuropsychological instruments to be used with young children. The NEPSY was designed to measure neuropsychological functions of children in the five functional domains of: Attention/Executive Functions, Language, Sensorimotor Functions, Visuospatial Processing, and Memory and Learning (Korkman et al., 1998). The NESPY-II includes these five functional domains, as well as an additional Social Perception domain.

Although Korkman and colleagues (1998, 2007b) purport that the five functional domains of the NEPSY and the six domains of the NEPSY-II do not develop in isolation but rather work in concert together, the domains are considered to be useful sources for specific groupings of cognitive function. The Attention/Executive Functions domain, of both the NEPSY and NEPSY-II, is considered by test authors to be a valid measure of abilities of inhibition, monitoring and self-regulation, vigilance, selective and sustained attention, flexibility in responding, nonverbal problem solving and figural fluency (Korkman et al.).

The subtests that compose the Attention/Executive Functions domain of the NEPSY are Tower, Auditory Attention and Response Set, Visual Attention, Statue, Design Fluency, and Knock and Tap. In the Tower subtest the child moves three colored balls situated on three pegs in order to match a picture in a certain number of moves. This subtest is assumed to assess the executive functions of planning, monitoring, self-regulation, and problem solving. In the Auditory Attention and Response subtest the child is first asked to select certain colors when they hear the color name through an auditory recording. The child is then asked to change their mental set in order to respond to opposite color stimuli. They must also avoid selecting other specific colors when the color name is heard. Auditory Attention and Response Set is thought to measure vigilance, selective auditory attention, inhibition, mental flexibility, and the maintenance of a complex mental set. In the Visual Attention subtest the child is required to scan a specific array of pictures in order to mark targets as quickly and accurately as possible. This subtest assesses the speed and accuracy with which a child can perform a visual scan and locate a target. In the Statue subtest the child is required to stand still for a specified amount of time while withholding a response to distracters. This subtest measures inhibition as well as motor persistence. The Design Fluency subtest requires the child to create as many unique designs as possible by connecting two or more dots in either a structured or unstructured array. This test is thought to measure a child's ability to generate novel designs quickly. In the Knock and Tap subtest a child learns a specific pattern of response and then must maintain that mental set while inhibiting the impulse to portray the examiner's action. The child is then required to shift their mental set, maintaining and regulating additional responses to conflicting stimuli. This test assesses self-regulation and inhibition of response to conflicting verbal and auditory information.

The subtests that compose the Attention/ Executive Functions domain of the NEPSY-II are Animal Sorting, Auditory Attention and Response Set, Clocks, Statue, Design Fluency, and Inhibition. Three subtests from the NEPSY (e.g., Tower, Knock and Tap, and Visual Attention) were replaced with new subtests in the NEPSY-II (e.g., Animal Sorting, Clocks, and Inhibition). Animal Sorting is a card sorting task that requires children to form basic concepts, categorize, and shift from one category to another fluently. Clocks assesses planning, organization, visual perceptual skills, visual spatial skills, and the concept of time by instructing children to draw times on analog clocks. The Inhibition subtest requires the examinee to look at a series of shapes or arrows and quickly label the shape or direction, or inhibit typical responding by providing an alternative response. Inhibition measures the ability to suppress automatic responses in order to employ mental flexibility between response types (Korkman et al, 2007a).

A limitation of the Attention/Executive Functions domain of the NEPSY and the NEPSY-II is an inadequate amount of empirical research on the subcomponents of executive functions and the factors that they purport to measure. The selection of subtests for the Attention/ Executive Functions Domain of the NEPSY was built on studies identifying convergence in the executive functioning factors of planning (Levin et al., 1991), speed, and fluency (Mirsky, Anthony, Duncan, Ahearn, & Kellan, 1991; Welsh et al., 1991); however, research supporting that the subtests purported to measure these factors are adequate identifiers of executive functioning skills has yet to be presented. Even less empirical research examining the Attention/ Executive Functions domain of the NEPSY-II is currently available. Research that supports the efficacy of the NEPSY and

NEPSY-II with clinical populations of children (e.g., Autism Spectrum Disorders, Specific Learning Disorders, ADHD, language impaired, intellectually disabled, deaf or hard of hearing, emotionally disabled, and spina bifida) has been examined (Hooper, Poon, Marcus, & Fine, 2006; Joseph, McGrath, & Tager-Flusberg, 2005; Riddle, Morton, Sampson, Vachha, & Adams, 2005). Research has demonstrated that a group of individuals identified with a learning disability in mathematics scored lower on the Auditory Attention and Response set subtest, individuals with Autism scored lower than controls on the Animal Sorting subtest, and a group of children with ADHD scored significantly lower on all subtests of the Attention and Executive domain of the NEPSY-II, except for Animal Sorting. These results are consistent across the NEPSY and NEPSY-II (Korkman et al., 2007b).

Delis-Kaplan Executive Function System (D-KEFS). The *Delis-Kaplan Executive Function System* (D-KEFS; Delis et al., 2001) presents a standardized assortment of existing and modified measures of executive functioning in children, adolescents, and adults between the ages of 8 and 89. The D-KEFS is the first grouping of executive functioning tests that is co-normed on a large stratified sample designated to assess the functions of mental flexibility, inhibition, problem solving, planning, impulse control, concept formation, abstract thinking, and verbal and spatial creativity (Homack, Lee, & Riccio, 2005). The D-KEFS was not developed on any single theoretical foundation, as the authors felt that because the knowledge of frontal lobe functioning is still in development, the association of the D-KEFS with a specific theoretical orientation would have been premature (Delis et al.). Rather, the D-KEFS consists of a variety of procedures that have been empirically demonstrated to be significant in the detection of executive dysfunction.

The D-KEFS is a compilation of nine stand-alone tests that are each individually aimed at assessing relevant verbal and nonverbal executive functions. Eight of these tests are administered to children, including the Word Context Test, Sorting Test, Twenty Questions Test, Tower Test, Color-Word Interference Test, Verbal Fluency Test, Design Fluency Test, and Trail Making Test. In order to avoid a lengthy, extended evaluation, the tests are arranged so that clinicians can choose the most appropriate subparts to administer, based on individual needs (Baron, 2004). Several of the tests contain multiple conditions and levels to the tasks. A brief review of each of the D-KEFS tests is discussed subsequently.

The main purpose of the Word Context Test is for the examinee to identify the meaning of a made-up word based on clues in a sentence. For each mystery word the examinee is shown five sentences that aid in the decoding of the word. This test is designed to measure deductive reasoning, integration of information, hypothesis testing, and flexibility of thinking.

In developing the D-KEFS Sorting Test modifications were made to the original *California Card Sorting Test* (Delis, 1988). The Sorting Test is a collaboration of two testing conditions. In the Free Sorting condition, the examinee is asked to sort 6 cards into two groups of three, according to as many different concepts as possible and then has to describe the rule. In the Sort Recognition condition the examiner sorts the cards into two groups and the examinee has to identify the overarching rule. This test is purported to measure the individual's ability to initiate problem solving skills both verbally and nonverbally.

In the Twenty Questions Test the examinee is presented with a page displaying the pictures of 30 ubiquitous objects, and then is required to ask the fewest possible number of questions in order to identify the unknown object. The executive functions measured by this test are abstract thinking and the identification of categories.

The Tower Test requires the examinee to move a series of disks across three pegs in order to match a picture in the shortest amount of moves. The examinee is instructed to only move one disk at a time and to never place a large disk over a smaller one. The Tower Test measures spatial planning, rule learning, inhibition of impulsiveness, inhibition of perseverative responding, and maintaining cognitive set.

The Color-Word Interference Test is a Stroop task that consists of four conditions that measure inhibition and cognitive flexibility. The first two conditions require participants to name color patches and read words that denote colors printed in black. The next two conditions require participants to inhibit reading colors by reading a discordant ink color, and to switch back and forth between naming the color of the word and the dissonant colored word.

The Verbal Fluency Test consists of the three components of Letter Fluency, Category Fluency, and Category Switching. In Letter Fluency, the examinee is asked to say words that begin with a particular letter as quickly as possible. In Category Fluency the participant is asked to verbalize words of a specific semantic category as quickly as possible. The last condition of Category Switching requires the examinee to alternate saying words of two different categories as quickly as possible. This test is purported to measure the examinee's word fluency skills.

The test of Design Fluency taps the functions of response inhibition, design fluency, and cognitive flexibility through the use of three conditions. The first condition requires the examinee to draw as many designs as possible in a series of boxes containing five dots. In condition two the examinee is required to connect unfilled dots and inhibit connecting any filled dots. In condition three the examinee has to switch between drawing designs in filled and unfilled dots.

The Trail Making Test consists of five conditions which require the examinee to sequentially connect letters and numbers according to varying rules. The participants are required to connect numbers in order as quickly as possible, connect letters in order as quickly as possible, and to switch between connecting letters and numbers. This test measures flexibility of thinking, visual scanning, number sequencing, letter sequencing, and motor speed.

Although various validity studies indicate that the individual measures of the D-KEFS are fairly accurate in distinguishing the clinical groups of fetal alcohol exposure, schizophrenia, chronic alcoholism, Alzheimer's disease, and Huntington's disease (Delis et al., 2001; Homack et al., 2005), the usefulness of the D-KEFS with clinical populations of children is largely unknown. In one study of clinical validity, Wodka and colleagues (2008) examined the performance of children with ADHD on four subtests of the D-KEFS (Trail Making, Verbal Fluency, Color-Word Interference, and Tower). The results indicated that even though children with ADHD performed in the average range on all four tasks, they exhibited significant differences from controls on Color-Word Interference and Tower when using summary measures. Additionally, specificity was not apparent in determining differences between the three ADHD subtype groups, which is consistent with similar ADHD research (Pasini, Paloscia, Alessandrelli, Porfirio, & Curatolo, 2007).

Evidence of validity in the D-KEFS was also tested and reported via correlations of the D-KEFS with the *California Verbal Learning Test-II* (CVLT-II; Delis, Kramer, Kaplan, & Ober, 2000) and *Wisconsin Card Sorting Test* (WCST; Heaton, 1981), in addition to findings within clinical populations. Convergent validity was determined by correlations between the D-KEFS, the *California Card Sorting Test* (Delis, 1988), and the WCST, where moderate correlations were found. Evidence of discriminate validity was determined by a lack of correlation between the D-KEFS tests and the CVLT-II, which is a verbal memory test.

In a recent factor analytic study, Latzman and Markon (2010) examined the factor structure of the D-KEFS from data collected from the standardization sample, as well as in a group of 11- to 16-year-old males. A series of exploratory factor analyses revealed that the D-KEFS subtests indicated a three factor solution across both samples. The three factors that were identified by test authors were labeled conceptual flexibility, monitoring, and inhibition. The analysis revealed that data did not appear to fit a universal, one-factor model of executive functioning (Latzman & Markon).

Validity Issues across the WJ III COG, NEPSY, and D-KEFS

Validity refers to the ability to accurately determine if a test is measuring what it is intended to measure. Validity is an essential piece of assessment, as without it accurate conclusions about a test's results would be unattainable, if not erroneous. Identifying the validity of a test involves acquiring information about the test's construct validity, content validity, and criterion-related validity.

Construct validity denotes that a test is measuring the construct that it is professed to measure (Mitchell & Jolley, 2004). Variants of construct validity are convergent validity (i.e., the degree to which two theoretically similar measures correlate) and divergent validity (i.e., the degree to which two theoretically divergent measures differ). With regards to psychological, cognitive, and neuropsychological assessment tools, convergent construct validity can be determined by the degree to which a specific measure accurately predicts performance of the targeted clinical groups of individuals (e.g., a depression measure that accurately predicts individuals with depressive disorders). Divergent construct validity can be established by the degree to which a specific measure does not correlate with the performance of an untargeted group of individuals (e.g., a depression measure that does not significantly predict the performance of individuals with verbal learning deficits).

The evidence of construct validity is further determined by the extent of content and criterion-related validity in a test measure. Content validity refers to the extent to which the test contains a fair sampling of the larger body of knowledge and relevant dimensions of the field of research (e.g., a measure that assesses all components of depression would have good content validity). In order to obtain substantial content validity a test must include samplings from every dimension of the measured construct and have a large sample from each of those dimensions (Mitchell & Jolley, 2004).

The criterion-related validity of a test is evident when the test demonstrates its effectiveness in determining or predicting the indicators of a construct. Criterion-related validity can be separated into two types of validity, concurrent and predictive. Concurrent validity studies investigate the correlation of performance on two or more measures obtained simultaneously. Concurrent validity can also be categorized by convergent and discriminate validity. Convergent concurrent validity entails substantiating that measures that are purported to identify similar constructs highly correlate with one another when administered simultaneously. Discriminate concurrent validity is intended to demonstrate the weaker relationship between two variables that are not intended measure the same construct. In contrast, predictive validity is aimed at determining that the performance on one test can estimate future performance on that variable (Mitchell & Jolley, 2004).

As previously discussed, the validity of executive functioning tasks across the WJ III COG, NEPSY, and D-KEFS has not yet been confirmed in the literature. Specifically, research has yet to confirm the construct, content, and criterion-related validity of these tasks. However, two studies have focused on the concurrent construct validity of executive functioning tasks in two of the three test batteries. Carper (2003) examined the subtests of the Executive Processes Cluster of the WJ III COG (Concept Formation, Planning, and Pair Cancellation) and the Tower and Design Fluency subtests of the NEPSY in a sample of typically developing children ($N = 60$). Results from this study indicated that Design Fluency from the NEPSY correlated with the WJ III COG Executive Processes Cluster ($r = .78$, $p < .01$) but not with any of the individual subtests of the Executive Processes Cluster. Furthermore, Tower from the NEPSY demonstrated no significant correlation at all. Interestingly, when the sample was broken down by age, Design Fluency only correlated with the Executive Processes Cluster in 12 year olds, but not in children age 10 or 11. When utilizing a Fisher's R to Z transformation to normalize the sampling distribution, no significant age differences were found (Carper).

Floyd and colleagues (2006) identified the relationships between the Executive Processes, Broad Attention, Working Memory, and Cognitive Fluency clinical clusters of the WJ III COG and the D-KEFS in a sample of children ($N = 92$) and adults ($N = 100$). Results from this study indicate a pattern of significant positive relationships between the WJ III COG clinical clusters and the subtests of the D-KEFS in both samples. Furthermore, of the four WJ III COG clinical clusters, the executive processes cluster evidenced the strongest and steadiest relationship with the D-KEFS tasks. However, it should be noted that the D-KEFS Tower test was not significantly correlated with any of the WJ III COG executive functioning clinical clusters in children; however, it was significantly correlated with the executive processes clinical cluster of the WJ III COG in the sample of adults ($r = .25$, $p < .01$). Although significant correlations were evident in both age samples; overall, stronger correlations between the two test batteries were apparent in the sample of children. Differences in the intensity of the correlations between the two samples suggest that specific strategies and abilities used to solve executive functioning tasks may differ depending on age (Floyd et al.). It should be noted that this study did not identify the relationships between specific executive functioning subtests of the WJ III COG with D-KEFS tasks.

Although the two aforementioned studies have contributed to the understanding of the utility of these tasks in children, more research is needed. Specifically, measuring the validity of these executive functioning tasks in a clinical group of children will further demonstrate if the selected subtests are good predictors of executive dysfunction. Identification of specific constructs of executive functioning by utilizing test validity will help determine the boundaries and developmental significance of executive functioning in children and adolescents.

Rationale and Purpose of Current Study

This chapter has demonstrated that although research on executive functions is abundant, research is greatly varied, leading to a vast misunderstanding of the structure and functions of the executive system. Because the boundaries of executive functioning are still unknown, yet the cognitive components that they represent are so important to human functioning, the need for understanding these processes is crucial. Even the physiological structure and localization of executive functioning skills is debated and greatly varied in the literature. Additionally, because the executive system is thought to play such a critical role in human functioning, many practitioners in the field of psychology, neuropsychology, and neuroscience often assess for neurocognitive strengths and weaknesses related to the executive system. As discussed in this chapter, a plethora of stand-alone executive functioning measures, cognitive batteries, and neuropsychological batteries are available for practitioners' use. However, research has demonstrated that these assessments often fail to demonstrate specific, valid measurements of executive functioning abilities. Furthermore, research on the construct validity of executive functioning scales of some of the larger cognitive and neuropsychological assessments, such as the WJ III COG, the NEPSY, and the D-KEFS, is limited in the literature. This is problematic since results from these assessment tools, even when validity is unknown, often lead to clinical diagnoses and interventions strategies for future goals. Determining reliable and valid measures that accurately reflect the structure of executive functioning will support more accurate neurocognitive evaluations.

Lack of research regarding valid and appropriate executive functioning measures is even more significant when these assessment tools are used with children. Very few executive functioning measures were designed specifically for children, in fact most assessments tools used with children are modifications of adult scales. Specifically examining executive functioning in children is imperative to the understanding of what processes of executive functions manifest at different developmental levels. Further research in this area will continue to bring information about related childhood disorders to the surface. Utilizing pure and valid measures of executive functions, developed specifically for children, will aid in the developmental discovery of the significance of executive dysfunction in children. Furthermore, continued research and the use of more valid assessment measures will ensure more accurate clinical diagnoses, educational placements, and intervention strategies for children with executive functioning difficulties.

Another issue that affects the inconsistency of executive functioning literature is multiple competing theories and models of executive functioning. Examining the validity of theories of executive functioning is crucial, as many of these theories have been utilized as the framework for executive functioning assessment tools. Determining the underlying factor structure of various executive functioning measures, and their fit with proposed theories of executive functioning is an important next step in research. If future fit statistics and factor analytic studies reveal adequate consistency and validity among different measures, and acceptable fit to proposed executive functioning theories, neuroimaging studies can then be utilized to further address the localization of executive functioning processes. Pinpointing how exactly executive functioning and executive dysfunction impact child development is necessary for the most applicable intervention strategies to be used for individuals with impairment to executive functioning processes.

The purpose of the current study was to examine some of these issues by first determining the degree of correlation among executive functioning subtests of three cognitive and neuropsychological assessments, the WJ III COG, the NEPSY, and the D-KEFS. This will help to determine the concurrent validity of these executive functioning tasks. These three assessment batteries were chosen for the current study due to the theoretical or foundational basis of each test. The WJ III COG has been selected for the current study because it is theoretically grounded on the CHC theory of cognitive abilities. The CHC theory is considered to be one of the most comprehensive available cognitive theories that is also applicable across multiple test batteries (McGrew, 2005). The NEPSY is included in the current study do to its assimilated executive functioning and attention domain. This is important, as recent research has demonstrated that these constructs overlap significantly (Baron, 2004; Boone et al., 1998; Brocki & Bohlin, 2004; Busch et al., 2005; Gioia et al., 2001; Kane & Engle, 2002; Rothbart et al., 2007). Lastly, the D-KEFS has been selected because it was designed to only measure constructs of executive functioning (Delis et al., 2001).

The target population of the current study is children of a mixed clinical sample. As already discussed, determining the validity of these assessments when used with children is an important area of research that is currently lacking in the field. Analyzing these constructs within a clinical group of children is advantageous since these are the children that are typically administered measures of executive functioning.

Additionally, the fit of various theories of executive functioning with the three assessment batteries is being examined in the current study in order to determine the efficacy of using these theories to compose assessment tools. Specifically, Anderson and colleagues' (2002) model of executive functioning, the CHC theory (McGrew, 2005), and the SNP conceptual model (Miller, 2007, 2010) have been selected. Anderson and colleagues' model has been selected because it focuses specifically on executive functioning, while also encompassing a wide range of executive functioning subcomponents. The CHC theory was chosen due to its influence on cross-battery assessment, which is an important component of neuropsychological evaluations. The SNP conceptual model was selected because of its specific focus on neuropsychological assessment in children. Furthermore, a simple, one-factor executive functioning model was selected to be used as a comparison tool in contrast to the more advanced models.

CHAPTER III

METHOD

The purpose of this chapter is to propose a research study based upon key issues concerning the concurrent validity of executive functioning tasks in the *Woodcock Johnson III Tests of Cognitive Abilities* (WJ III COG; Woodcock et al., 2001c), the *Delis Kaplan Executive Function System (*D-KEFS; Delis et al., 2001), and *The NEPSY, A Developmental Neuropsychological Assessment* (NEPSY; Korkman et al., 1998). Issues related to the fit of executive functioning theories and models to the aforementioned tasks of executive functioning will also be discussed. Information about the current participants, data collection, procedures, measures, and data analysis will be outlined. Additionally, the research question, models addressed in the current study, and methodological issues will be identified.

Research Participants

The data utilized in the study was archival and was collected from case studies submitted as part of course completion requirements for the KIDS, Inc.'s School Neuropsychology Post-Graduate Certification Program. Individuals attempting to gain certification from the KIDS, Inc.'s School Neuropsychology Post-Graduate Certification Program were required to submit three comprehensive case studies completed on children with known or suspected neurocognitive strengths and weaknesses. Therefore, the data includes information from children between four to eighteen years of age, with a wide range of clinical diagnoses. Diagnoses within the sample include attention deficit hyperactivity disorder, learning disability, autism spectrum disorders, traumatic brain injury, and speech and language disorders. Additionally, report date, gender, ethnicity, languages, age, and grade were also collected for the current study.

Due to the selected assessment instruments, the study utilized information from a mixed clinical sample of children between the ages of eight and twelve. This age range was selected because of the subtest floor of the D-KEFS and the subtest ceiling of the NEPSY. In addition, this specific age range is appropriate to study due to the developmental trajectory of executive functioning skills in children this age. Each test administration and case report was required to meet certain standardization requirements based on training criteria set forth by the KIDS, Inc.'s School Neuropsychology Post-Graduate Certification Program. However, each case report contained unique assessment batteries based on the individual referral question. Therefore, cases were chosen based on availability of WJ III COG, NEPSY, and D-KEFS test scores. The total sample size was 321 cases. Issues related to the effect of this sample size on statistical power will be discussed later in this chapter.

Procedure

Each case report was obtained from submission for course completion requirements for the KIDS, Inc.'s School Neuropsychology Post-Graduate Certification Program. Any case report with an attached consent form stating that the data should not be used for research purposes was excluded from the study. The archival data was coded in order to ensure confidentiality of data. The data was maintained under the code and separated from the actual case file. Demographic data was culled from the case file in order to distinguish age, sex, grade, ethnicity, language, and diagnostic category.

Due to the use of archival data, and an inability to manipulate independent variables, a non-experimental research design was used in the current study. Overall test scores and composite cluster scores for the WJ III COG, NEPSY, and D-KEFS were recorded as well as the scores from executive functioning, attention, and working memory subtests. As discussed in the preceding chapter, executive functioning, attention, and working memory have all been implicated as playing a role in the executive system. Subtests were designated as measuring executive functioning, attention, or working memory based on the subtest descriptions proposed by the test authors.

Measures

Woodcock Johnson III Tests of Cognitive Abilities (WJ III COG)

The *Woodcock Johnson III Tests of Cognitive Abilities* (WJ III COG; Woodcock et al., 2001c) is a comprehensive and individually administered set of twenty norm-referenced tests for measuring intellectual abilities. The twenty subtests of the WJ III COG were each designed to measure different, selective features of cognitive ability. Test authors assert that only the most relevant tests should be administered during each assessment (Schrank, Flanagan, Woodcock, & Mascolo, 2002).

The WJ III COG was co-normed with the *Woodcock-Johnson III Tests of Achievement* (WJ III ACH; Woodcock, McGrew, & Mather, 2001b), which includes 22 oral language and academic achievement tests. The WJ III ACH is available in two parallel test forms that can be used to monitor academic changes while reducing implications of practice effects. The WJ III ACH is also available in a brief version. Together, the WJ III COG and the WJ III ACH comprise the *Woodcock-Johnson III test battery* (WJ III; Woodcock, McGrew, & Mather, 2001a). In addition, eleven supplemental cognitive measures are available through the *Woodcock-Johnson III Diagnostic Supplement to the Tests of Cognitive Abilities* (Diagnostic Supplement; Woodcock, McGrew, Mather, & Schrank, 2003).

Development. The WJ III COG denotes the third generation of cognitive tests that were originally developed in 1973 and included in the *Woodcock-Johnson Psycho-Educational Battery* (WJPEB; Woodcock & Johnson, 1977; Schrank et al., 2002). The *Woodcock-Johnson Psychoeducational Battery – Revised: Tests of Cognitive Ability* (WJ-R COG; Woodcock & Johnson, 1989) was later developed and theoretically grounded in the *Gf-Gc* theory (Horn, 1991) of cognition. The WJ III COG was subsequently developed to correspond with the integrated cognitive theory that came to be known as the Cattell-Horn-Carroll theory of cognitive abilities (CHC theory; McGrew, 2005, McGrew & Woodcock, 2001).

CHC theory asserts that a hierarchical, three-tier model of cognitive functioning exists (McGrew, 2005, McGrew & Woodcock, 2001). CHC theory suggests that there is an overarching "*g*" factor that describes overall intelligence. Subsumed by the *g* factor are several broad ability factors that describe categories of intelligence, such as short-term memory and processing speed. The lowest tier contains multiple narrow abilities that describe even more specific aspects of intelligence (McGrew; McGrew & Woodcock).

Although many features of the WJ-R COG were retained during the revision and renorming process, several important changes also occurred. First of all, the factor structure of the WJ III COG includes two to three tests that each measures a broad CHC ability. Furthermore, each test subsumed within a broad ability cluster measures a different and unique narrow ability. This allows for qualitative interpretation of differences found within the same broad ability. In addition, oral language tests were moved to the WJ III ACH, and eight new tests of working memory, planning, naming speed, and attention were added to the cognitive battery. Additionally, five clinical clusters were added to improve diagnostic usage and clinical sensitivity. Furthermore, many of the tests were redesigned to be more suitable for use with bilingual subjects and to include a greater variety of norms (Mather & Woodcock, 2001).

In 2005, the WJ III normative data was recalculated and renormed based on updated U.S. Census statistics, which resulted in the development of the *Woodcock-Johnson III Normative Update* (WJ III NU; McGrew et al., 2007). It is important to note that only the normative data was updated, and thus, WJ III COG and WJ III ACH content remained intact after renorming. In the subsequent discussion of test content, validity, and reliability, the distinction between the WJ III COG and the WJ III COG NU will only be made when normative data is specifically discussed. In all other instances, when content between the WJ III COG and the WJ III COG NU is indistinguishable, the assessment will be referred to as the WJ III COG.

Content. Due to its foundation in CHC theory, the WJ III COG is arranged in three hierarchical levels of performance measurement that include general ability, broad ability, and narrow ability scores. General ability is measured by the comprehensive General Intellectual Ability (GIA) score. Although the WJ III COG includes a general ability score, the primary focus of the WJ III COG is in the measurement of the broad CHC factor scores (Schrank et al., 2002). The seven broad cluster abilities included in the WJ III COG are Comprehension-Knowledge (*Gc*), Long-Term Retrieval (*Glr*), Visual-Spatial Thinking (*Gv*), Auditory Processing (*Ga*), Fluid Reasoning (*Gf*), Processing Speed (*Gs*), and Short-Term Memory (*Gsm*). Narrow abilities are listed by each subtest and serve as additional information for qualitative interpretation (Schrank et al.).

In addition to the GIA and broad CHC factor scores, the WJ III COG also includes general clusters that split cognitive abilities into broad aspects of cognitive performance, namely Acquired Knowledge, Thinking Ability, and Cognitive Efficiency. Several clinical clusters may also be obtained from combinations of standard and supplementary subtests within the WJ III COG that are useful in the diagnostic or clinical setting. These clusters include Phonemic Awareness, Working Memory, Broad Attention, Cognitive Fluency, and Executive Processes (Mather & Woodcock, 2001). Scores from subtests within the Executive Processes, Broad Attention, and Working Memory clinical clusters will be included in the current study (see Table 1 for a list of subtests, broad abilities, and narrow abilities found within these clinical clusters). The three selected clinical clusters are all noted to measure aspects of executive functioning by the test authors (Mather & Woodcock).

Table 1

Woodcock Johnson III Tests of Cognitive Abilities, Broad and Narrow Abilities

Name of Subtest	CHC Broad Abilities	CHC Narrow Abilities
Executive Processes Cluster		
Concept Formation	Fluid Reasoning (*Gf*)	*Induction*
Planning	Fluid Reasoning (*Gf*) Visual-Spatial Thinking (*Gv*)	*Spatial spanning* *General sequential reasoning*
Pair Cancellation	Processing Speed (*Gs*)	*Attention and concentration*
Broad Attention		
Numbers Reversed	Short-Term Memory (*Gsm*)	*Working Memory*
Auditory Working Memory	Short-Term Memory (*Gsm*)	*Working Memory*
Auditory Attention	Auditory Processing (*Ga*)	*Speech-sound discrimination* *Resistance to auditory stimulus distortion*
Pair Cancellation	Described in Executive Processes Cluster	
Working Memory		
Numbers Reversed	Described in Broad Attention Cluster	
Auditory Working Memory	Described in Broad Attention Cluster	

The cluster of Executive Processes in the WJ III COG incorporates three aspects of executive functioning, namely, strategic planning, interference control/ inhibition, and

mental flexibility. The three tests that form this cluster are Concept Formation, Planning, and Pair Cancellation. The subtest of Concept Formation is a controlled learning task that requires rule formation after a specific set of stimuli has been presented. There is no memory component to this subtest as there is a stimulus key presented throughout. This subtest measures the controlled ability to alter one's mental set (Mather & Woodcock, 2001). This task also utilizes induction, categorical reasoning, logic, concept shifting, and categorization. Within the Planning subtest the examinee is required to employ forethought and mental restriction in order to trace a given stimulus without removing the pencil from the piece of paper or retracing any lines. The narrow abilities that the Planning subtest requires are spatial scanning and general sequential reasoning.
This task also requires planning and forethought. The Pair Cancellation subtest is a timed test that requires the examinee to locate and mark a specific pattern of repeating objects on a provided stimulus card. It requires the capacity to stay on task in a vigilant manner and requires sustained attention, concentration, recognition, monitoring, and interference control (Mather & Woodcock).

The Broad Attention cluster assesses a global overview of attention (Mather & Woodcock, 2001). The four subtests within the Broad Attention cluster, Numbers Reversed, Auditory Working Memory, Auditory Attention, and Pair Cancellation, each measure different aspects of attention. In Numbers Reversed the subject mentally holds a series of numbers while reversing the sequence. It requires attentional capacity, working memory, and transformation. In Auditory Working Memory, the subject listens to a series of numbers and words, and then repeats the words in sequential order, followed by the numbers in numerical order. Auditory Working Memory necessitates divided attention, working memory, reorganization, sorting, and sequencing. In Auditory Attention, the subject is asked to listen to phonetically similar words (i.e., bee, knee, sea), while presented with increasingly intense background noise. Subjects are then required to point to a picture that represents the word they heard. This task entails selective attention, speech-sound discrimination, and inhibition of extraneous auditory stimuli. As already discussed with the executive processes cluster, Pair Cancellation requires sustained attention. The Working Memory cluster includes two subtests from the Broad Attention Cluster, Numbers Reversed and Auditory Working Memory. The Working Memory clinical cluster does not include any unique tests, but has been defined by test authors to be a good indicator of executive functions (Mather & Woodcock).

Scores from the six identified WJ III COG tests are included in the current study, namely, Concept Formation, Planning, Pair Cancellation, Numbers Reversed, Auditory Working Memory, and Auditory Attention. Although subtests from the Executive Processes, Broad Attention, and Working Memory clinical clusters will be utilized in structured factor modeling, the broad cluster scores will not be employed in the study. Furthermore, broad CHC ability scores will not be directly utilized, but will be taken into account for theoretical modeling.

Standardization and norm development. Normative data for the WJ III were originally based on a sample of 8,818 individuals, aged two to over ninety, who were given both the WJ III COG and the WJ III ACH. Individuals included in the sample closely resembled the demographic characteristics of the United States (U.S.) 2000 census (Mather & Woodcock, 2001) and data was collected from over 100 geographically diverse regions. Subjects were randomly selected through a three stage stratified sampling design that controlled for census region, community size, sex, race, type of school, type of college or university, education of adults, occupational status of adults, and occupations of adults in the labor force. Normative data was collected by teams of trained examiners under direct supervision (McGrew & Woodcock, 2001).

In 2005, the WJ III normative data was recalculated based on updated U.S. Census statistics, which resulted in the WJ III NU (McGrew et al., 2007). The WJ III NU sample included 8,782 subjects from ages 12 months to 80 years and older. Subjects were randomly selected through a three-stage stratified sampling design and controlled for census region, community size, sex, race, type of school or university, education, occupational status, and nation of birth. Individuals in the updated normative sample were administered the WJ III COG, the Diagnostic Supplement, and the WJ III ACH. Normative data was collected by teams of trained examiners under direct supervision. Additionally, the WJ III NU norms replaced the original WJ III norms in the updated computer scoring program (McGrew et al.). Participants included in the current study may be compared to year 2000 or 2005 norms, based on when they were administered the WJ III.

Reliability. Internal consistency using split-half reliability procedures as well as test-retest reliability were assessed for the subtest and cluster scores of the WJ III COG and WJ III COG NU. For both normative versions, split-half procedures were calculated for all but the speeded tests and tests with multiple-point scoring systems. Split-half calculations were conducted by separating odd and even test items and comparing the internal consistency of the two halves. Split-half calcluations were corrected for length of test by using the Spearman-Brown correction formula (Crocker & Algina, 1986; Mather & Woodcock, 2001; McGrew et al., 2007).

Reliabilities for the WJ III COG and WJ III COG NU speeded tests (Visual Matching, Retrieval Fluency, Decision Speed, Rapid Picture Naming, and Pair Cancellation) and tests with multiple-point scored items (Spatial Relations, Retrieval Fluency, Picture Recognition, and Planning) were calculated using Rasch analysis procedures (Mather & Woodcock, 2001; McGrew et al., 2007, Rasch, 1980). This involved assessing speeded tests through a test-retest procedure where subjects were administered the same tests two times. For both normative versions, tests were presented in a counter-balanced order at retest in order to reduce test familiarity confounds. The retest interval for speeded tests was originally set at one day to minimize changes in subjects' states or traits. Additional test-retest reliability studies were conducted for fifteen cognitive and academic tests, with the retest interval ranging from less than one year to ten years (Mather & Woodcock; McGrew et al.).

Table 2 reports the WJ III COG median reliability coefficients (r_{11}) and the standard errors of measurement (SEMs) that were obtained using the split-half or test-retest procedures already discussed. Table 3 reports the median reliability coefficients (r_{11}) and the SEMs from the WJ III COG NU. The SEM represents how much error is associated with any test or cluster score. The tests and subtests with higher reliabilities generally have smaller SEMs. Only tests and cluster scores that are applicable to the current study are reported. Across both normative groups, all cluster reliability coefficients, except for Visual-Spatial Thinking and Short-Term Memory clusters, are .90 or higher, which is desirable for cluster scores (McGrew et al., 2007). Also across both normative groups, all subtest reliability coefficients, except for the Planning subtest, are .80 or higher, which is generally regarded as a desirable reliability coefficient for an individual test (McGrew et al.).

Table 2

Woodcock Johnson III Tests of Cognitive Abilities, Median Reliability Statistics

Cluster	Median r_{11}	Median SEM (SS)	Test	Median r_{11}	Median SEM (SS)
Executive Processes	.93	3.97	Concept Formation	.94	3.64
Broad Attention	.92	4.24	Planning	.74	7.65
Working Memory	.91	4.50	Pair Cancellation	.81	6.56
Fluid Reasoning (*Gf*)	.95	3.35	Numbers Reversed	.87	5.38
Visual-Spatial Thinking (*Gv*)	.81	6.54	Auditory Working Memory	.87	5.37
Short-Term Memory (*Gsm*)	.88	5.20	Auditory Attention	.88	5.21
Auditory Processing (*Ga*)	.91	4.50			

Table 3

Woodcock Johnson III Tests of Cognitive Abilities Normative Update, Median Reliability Statistics

Cluster	Median r_{11}	Median SEM (SS)	Test	Median r_{11}	Median SEM (SS)
Executive Processes	.96	3.00	Concept Formation	.94	3.67
Broad Attention	.94	3.67	Planning	.74	7.65
Working Memory	.91	4.50	Pair Cancellation	.96	2.92
Fluid Reasoning (*Gf*)	.95	3.35	Numbers Reversed	.87	5.41
Visual-Spatial Thinking (*Gv*)	.81	6.54	Auditory Working Memory	.87	5.41
Short-Term Memory (*Gsm*)	.88	5.20	Auditory Attention	.88	5.20
Auditory Processing (*Ga*)	.91	4.50			

Validity. The WJ III COG and WJ III COG NU are based on several types of validity evidence including content validity, internal validity, and external validity (Schrank et al., 2002; Schrank & Flanagan, 2003; McGrew et al., 2007). Content validity for the WJ III COG was developed using theory-based operational definitions of proposed abilities and constructs, based on CHC theory (Schrank & Flanagan). Tests within the WJ III COG were then revised or modeled after existing tests in cognitive and neuroscience research that parallel constructs found within CHC theory. Furthermore, content was reviewed by experts in CHC theory, who examined each test and cluster multiple times to determine fit with operational definitions and CHC theory constructs (Schrank & Flanagan). To further assess for content validity and potential sensitivity issues, additional expert reviewers evaluated items included in each test to determine and identify potential issues regarding gender, individuals with disabilities, and cultural or linguistic minority groups. Any items that had inadequate or biased content were eliminated or revised. Content validity was similarly examined with data from the normative group from the WJ III COG NU (McGrew et al.).

Internal structure was assessed by mapping the fit of test and cluster performance to CHC theory. Confirmatory factor analyses (CFA) were utilized to determine the match of WJ III COG tests to the proposed narrow abilities, broad abilities, and general intellectual ability outlined by CHC theory. When compared to multiple models of cognitive and intellectual ability, tests of the WJ III COG most closely fit with the three-tiered CHC model (Schrank & Flanagan, 2003).

External validity was assessed by comparing test performance to outside variables. Both convergent and divergent validity evidence exists for the WJ III COG and WJ III COG NU tests and clusters. Cross-sectional data demonstrates that divergent developmental growth curves are evident for each of the CHC broad and narrow abilities measured by the WJ III COG and WJ III COG NU. General performance on tests and clusters revealed score changes that parallel the developmental growth and decline of cognitive abilities throughout the life span, suggesting that each narrow and broad ability is unique (McGrew et al., 2007; Schrank & Flanagan, 2003).

External validity was further assessed by comparing WJ III COG scores to other intellectual tests scores that were taken by the same individuals. Convergent validity was apparent in several studies that compared the overall WJ III COG GIA scores to the full-scale scores of other cognitive batteries. Correlations with the *Wechsler Intelligence Scale for Children – Third Edition* (WISC-III; Wechsler, 1991) were reported to be .71 for the standard GIA score (GIA-Std.) and .76 for the extended GIA score (GIA-Ext). Correlations with the *Stanford-Binet Intelligence Scale, Fourth Edition* (SB-IV; Thorndike, Hagen, & Sattler, 1986) are reported as .76 for the GIA-Std. and .71 for the GIA-Ext. Overall correlations between the WJ III COG and other cognitive measures were moderate to high (McGrew et al., 2007).

Validity measures have also been determined for the WJ III COG clinical clusters. During the development of the Working Memory clinical cluster, experts operationally defined the CHC narrow ability of Working Memory (Schrank & Flanagan, 2003). Furthermore, McGrew and Woodcock (2001) provided CFAs that supported the grouping of tests within the Working Memory clinical cluster. External validity evidence for the Working Memory clinical cluster is supported by its correlations with working memory measures from related batteries (Shrank & Flanagan). Such batteries include the Working Memory Index of the *Wechsler Adult Intelligence Scale, Third Edition* (WAIS-III; Wechsler, 1997) and the Digit Span subtest of the WISC-III (Wechsler, 1991). In addition, the Working Memory cluster strongly correlates with components of reading, mathematics, and writing academic achievement found within the WJ III ACH. Similar content and external validity evidence has also been supported for the Broad Attention and Executive Processes clinical clusters; however, detailed descriptions were not available in test manuals (Schrank & Flanagan).

NEPSY, A Developmental Neuropsychological Assessment

The NEPSY, A Developmental Neuropsychological Assessment (NEPSY) was designed as an assessment tool for the neuropsychological development of preschool and school-age children between the ages of three and twelve (Korkman et al., 1998). The 27 subtests of the NEPSY were designed to assess fundamental and complex cognitive abilities that are essential for learning. The NEPSY assesses the five functional domains of Attention/ Executive Functions, Language, Sensorimotor, Visuospatial, and Memory and Learning.

The NEPSY was developed with four associated purposes in mind (Kemp et al., 2001). The first purpose was to create a valid and reliable neurocognitive instrument for children that was sensitive to strengths and weaknesses across the five identified functional domains. The second purpose was to contribute to the knowledge base of congenital or traumatic brain damage. The third purpose was for the NEPSY to be used in long term follow up of children with brain damage or dysfunction. Finally, the NEPSY was intended to study neuropsychological development in preschool and school-aged children (Kemp et al.).

In 2007 a revised version, *the NEPSY, A Developmental Neuropsychological Assessment, Second Edition* (NEPSY-II; Korkman et al., 2007b) became available. The NEPSY-II includes an extended age range (3-16), revised subtests, and an added functional domain (social perception); however, many of the subtests from the original

NEPSY remained unchanged in the NEPSY-II. The current study utilizes data from the NEPSY, as most of the archival data was collected prior to the NEPSY revision.

Development and standardization. The NEPSY was developed based upon the theoretical work of A.R. Luria, who theorized that human cognitive functions can be conceptualized within a framework of three functional but integrated units that he called "blocks." Lurian theory posits that the blocks are arranged according to complexity of cognitive functioning (Kemp et al., 2001; Luria, 1980). In addition to Lurian theory, the NEPSY was also based on contemporary advances in child neuropsychology and research (Kemp et al.).

The NEPSY was originally developed in Finland out of the need for adequate neuropsychological instruments for young children. The original Finish NEPSY was designed to measure neuropsychological functions of five- and six-year-old children in the five functional domains of: Attention/Executive Functions, Language, Sensorimotor Functions, Visuospatial Processing, and Memory and Learning (Korkman et al., 1998). The original version was later extended to include three- to nine-year-olds and encompassed a more psychometric approach to neuropsychological testing (Korkman et al.). The test was also made available in English, Swedish, and Danish versions. In 1987, the development of the American NEPSY began, with plans to include contemporary neuropsychological assessment research and maintain multicultural sensitivity (Kemp et al., 2001).

The American NEPSY was developed in three phases: the pilot phase, tryout phase, and standardization and validation phase. During the pilot phase (1987-1989) the

NEPSY was adapted for 3- to 10-year old children and new subtests were added (Kemp et al., 2001). Some of the added subtests were based on established neuropsychological assessments, such as Fingertip Tapping and Phonemic Fluency (Benton, Hamsher, Varney, & Spreen, 1983; Denckla, 1973). A 41 subtest pilot version of the NEPSY was administered to 160 American children in the Northeast who were randomly selected from urban and suburban settings and stratified for age, gender, and educational and socioeconomic background. Pilot administrations were also completed in Finland. Thereafter, some subtests were eliminated, modified, or created based on the results of the pilot studies, literature reviews, and clinical experiences. Additional pilot studies were conducted with the updated subtests. Furthermore, a bias and cultural sensitivity review was conducted, which resulted in the adjustment of items viewed as biased (Kemp et al.).

In 1991, the tryout stage of the NEPSY began, which involved the administration of 52 NEPSY subtests to 300 children across the United States (Kemp et al., 2001). Subtests were administered to children 2 to 12 years of age. The tryout sample was stratified by ethnicity, gender, geographic region, and parental education. A second bias review was conducted, and items viewed as biased or culturally insensitive were modified or deleted. After review of the tryout data, subtests with poor reliabilities, such as those for two-year-olds, were deleted. Subtest floor and ceiling problems, psychometric properties, and scoring procedures were reviewed and modified. Additional tryout administrations were conducted in America and Finland. After the final tryout data review, subtests for inclusion in the standardization of the NEPSY were selected (Kemp et al.).

The standardization and validation phase of the NEPSY was conducted from 1994 to 1996. Thirty-eight subtests were included in the standardization version of the NEPSY. The standardization sample totaled 1000 cases that included 100 children in each age group ranging from 3 through 12 years. The sample included 50 males and 50 females in each age group and was stratified for ethnicity, parent education, and geographic region. In addition, the proportion of ethnic groups was corrected according to the 1995 U.S. census and was maintained for each age group. After all of the national standardization and validation data was reviewed, the final selection of subtests was chosen for each of the five functional domains. The subtests within each domain were then classified as either core or expanded (Kemp et al., 2001).

Content. Korkman and colleagues (1998) suggest that the five functional domains of the NEPSY (Attention/Executive Functions, Language, Sensorimotor Functions, Visuospatial Processing, and Memory and Language) do not develop in isolation but rather work in concert together. However, each domain is considered an adequate and independent measure of its respective neuropsychological function. The Attention/Executive Functions domain is considered to be a valid measure of abilities of inhibition, monitoring and self-regulation, vigilance, selective and sustained attention, flexibility in responding, nonverbal problem solving and figural fluency (Korkman et al.). Only measures from the Attention/ Executive Functions domain will be subsequently discussed.

The subtests that compose the Attention/Executive Functions domain are Tower, Auditory Attention/ Response Set, Visual Attention, Statue, Design Fluency, and Knock and Tap. Due to restricted age ranges and the use of archival data collection, the current study utilizes data collected from the following subtests: Tower, Auditory Attention/ Response Set, and Visual Attention. In the Tower subtest the child moves three colored balls situated on three pegs in order to match a picture in a certain number of moves. This subtest is assumed to assess the executive functions of planning, monitoring, self-regulation, and problem solving. In Auditory Attention the child is asked to select certain colors when they hear the color name through an auditory recording. They must also avoid selecting other specific colors when the color name is heard. This subtest measures simple selective and sustained auditory attention, as well as inhibition. Response Set is administered directly following Auditory Attention. In this test the child is asked to change their mental set in order to respond to opposite color stimuli. Response Set is thought to measure vigilance, selective auditory attention, mental flexibility, and the maintenance of a complex mental set. In the Visual Attention subtest the child is required to scan a specific array of pictures in order to mark targets as quickly and accurately as possible. This subtest assesses sustained attention and the speed and accuracy with which a child can perform a visual scan.

Reliability. Reliabilities were calculated for the Attention and Executive Functions domain as well as for each subtest using split-half reliabilities, test-retest reliabilities, or generalizability coefficients (Korkman et al., 1998). Split-half reliabilities were conducted on subtests that could be divided in half and were of similar length. The correlation coefficient derived from the split-half procedure was corrected for length using the Spearman-Brown formula (Crocker & Algina, 1986). Test-retest reliability was reported for subtests that could not be split in two parallel forms due to a multipart scoring or timing component. Generalizability coefficients were conducted on subtests that have multiple sources of potential error, such as a total score based on accuracy and speed.

Reliabilities were calculated for each age level and then separated and averaged across two age groups: 3 to 4 and 5 to12. SEMs were also calculated for each test to determine the rate of error in each observed score. Some subtests are not administered to younger children (i.e., Tower, Auditory Attention/ Response Set) so correlation coefficients and SEMs are not available for 3- to 4- year-olds. Reliability coefficients and SEMs for subtests utilized in the current study are listed in Table 4. The overall reliability for the Attention/ Executive Functioning domain for 5-12-year-olds suggests a moderately high score. The reliability score was notably lower for 3-4 year olds, which is not surprising due to the inherent developmental variability of attention and executive functioning in this age range (Korkman et al., 1998). Reliability for the Tower and Auditory Attention/ Response Set subtests is of moderate strength, but lower for the Visual Attention subtest. Stability coefficients were also determined across a sample of 168 participants and were generally of moderate strength. The median stability coefficient for the Attention and Executive Functions domain is .68 for the age range of 3-4, and .67 for children aged 5-12.

Table 4

NESPY, Average Reliability Statistics

DOMAIN/ Subtest	Age 3-4 r_{11}	SEM	Age 5-12 r_{11}	SEM
ATTENTION/ EXECUTIVE	.70	8.22	.82	6.38
Tower	---	---	.82	1.32
Auditory Attention/ Response Set	---	---	.81	1.32
Visual Attention	.76	1.47	.71	1.62

Validity. Korkman et al. (1998) employed a variety of techniques aimed at increasing the content validity of the NEPSY, such as comparing test content to the performance of children with and without known neurodevelopmental and acquired disabilities. The NEPSY's content was further reviewed in order to retain core elements from Lurian theory while also including contemporary psychometric and neuropsychological research. Furthermore, groups of pediatric neuropsychologists and school psychologists were recruited to review the NEPSY for content and bias. After considering recommendations, as well as previous pilot studies, alterations were made to NEPSY subtests.

Construct validity was assessed by comparing intercorrelations between NEPSY subtests and domains. Overall, subtests within functional domains correlated more highly than subtests across the different domains. This provides evidence that the internal structure of the NEPSY is sound (Korkman et al., 1998).

The NEPSY was also evaluated and compared against other scales for children and neuropsychological tests, such as the WISC-III (Wechsler, 1991), the Bayley Scales of Infant Development – Second Edition (BSID-II, Bayley, 1993), the Benton Neuropsychological Tests (Benton et al., 1983), and the Children's Memory Scale (CMS, Cohen 1997) as an attempt to further examine validity issues (Ahmad & Warriner, 2001; Korkman et al., 1998). After comparisons with tests of similar domains, moderate correlations were found in the area of attention, executive functioning, language, visual-spatial processing, and memory.

Furthermore, in order to compare the diagnostic utility of the NEPSY, data was collected on multiple clinical groups with neurological or developmental disabilities. Clinical groups that were compared included individuals with Attention Deficit Hyperactivity Disorder, Learning Disabilities, Language Disorders, Pervasive Developmental Disorders, Fetal Alcohol Syndrome, Traumatic Brain Injury, or Hearing Impairments. The results of the clinical studies revealed that overall, the NEPSY is sensitive to divergent degrees of neurocognitive dysfunction. Strengths and weaknesses within domains were generally consistent with attributes of the respective disorder (Korkman et al., 1998).

Schmitt and Wodrich (2004) found that when an overall intelligence score is not controlled for, the neuropsychological validity of the NESPY appears to be adequate. However, when overall intelligence is accounted for, the validity of the NESPY appears to be questionable. Due to limited research in this domain, any assured estimate of validity regarding the NEPSY may be premature (Ahmad & Warriner, 2001; Schmitt & Wodrich).

Delis-Kaplan Executive Function System (D-KEFS)

The *Delis-Kaplan Executive Function System* (D-KEFS; Delis et al., 2001) consists of nine stand-alone measures of executive functioning for children and adults between the ages of 8 and 89. The D-KEFS is the first grouping of higher-order cognitive tests that is co-normed on a large stratified sample and designed to assess multiple aspects of executive functioning (Homack et al., 2005). The nine D-KEFS tests measure an extensive range of verbal and nonverbal executive functions. Each test can be administered individually or with other D-KEFS tests. Although some of the tests within the D-KEFS are unique, many are modified from research-supported and widely used tests of executive functioning, such as the Stroop (1935) and Tower tasks (Humes et al., 1997).

The D-KEFS was developed in order to fulfill several objectives. First of all, prior to the development of the D-KEFS, few contemporary, research-based assessment instruments for measurement of the frontal lobes were available. Additionally, the D-KEFS was created in order to provide a unique test consisting of a comprehensive assortment of multiple executive functioning tasks. Finally, the D-KEFS was developed to support a cognitive-process approach to assessment. This approach would allow the D-KEFS to provide multiple scores of several cognitive tasks, rather than only one general score.

Another important feature of the D-KEFS is that it allows for the isolation of fundamental skills that may affect performance on executive functioning tasks. For example, many of the D-KEFS tests include several conditions that measure basic skills, such as visual scanning and motor speed, as well as more advanced skills. This allows the examiner to rule out potential confounds that may have affected performance on executive functioning tasks (Delis et al., 2001).

Development and standardization. Due to continually emerging knowledge regarding frontal lobe functioning, the D-KEFS was not developed on any single theoretical foundation. Test authors felt the association of the D-KEFS with a specific theoretical orientation would have been premature (Delis et al., 2001). Rather, the D-KEFS consists of a variety of procedures that have been empirically demonstrated as significant in the detection of executive dysfunction.

Research and development for the D-KEFS took place over a span of ten years. Test development went through several stages, beginning with initial test design and pilot testing with examinees with and without documented brain injury. Results of the pilot study resulted in revisions of test design followed by a controlled tryout study. Data from the tryout study resulted in further revisions and completion of the final national standardization study.

The nationally representative standardization sample consisted of 1750 individuals, stratified based on the 2000 U.S. census for age, sex, ethnicity, years of education/years of parent education, and geographic location. For collection of standardization data, examiners were selected based on their experience with psychometric testing, and certification/ licensure. Examinees were excluded from the study if they endorsed any psychiatric or medical condition that could affect performance on cognitive tasks.

Content. The D-KEFS is a compilation of nine stand-alone tests that are each individually aimed at assessing relevant verbal and nonverbal executive functions. Eight of these tests are standardized for administration to children. In order to avoid a lengthy, extended evaluation, the tests are arranged so that clinicians can choose the most appropriate subparts to administer, based on individual needs (Baron, 2004; Delis et al., 2001). Several of the tests contain multiple conditions and levels to the tasks. The eight stand alone tests for children are the Word Context Test, Sorting Test, Twenty Questions Test, Tower Test, Color-Word Interference Test, Verbal Fluency Test, Design Fluency Test, and Trail Making Test.

The main purpose of the Word Context Test is for the examinee to identify the meaning of a made-up word based on clues in a sentence. For each mystery word the examinee is shown five sentences that aid in the decoding of the word. This test is designed to measure deductive reasoning, integration of information, hypothesis testing, and flexibility of thinking.

In developing the D-KEFS Sorting Test, modifications were made to the original *California Card Sorting Test* (Delis, 1988). The Sorting Test is a collaboration of two testing conditions. In the Free Sorting condition, the examinee is asked to sort six cards into two groups of three, according to as many different concepts that they can self-generate. They then have to describe the rule for each sort. In the Sort Recognition condition the examiner sorts the cards into two groups and the examinee has to identify the overlying rule. This test measures an individual's ability to initiate problem solving skills both verbally and nonverbally, mental flexibility, and problem solving.

In the Twenty Questions Test the examinee is presented with a page displaying the pictures of thirty objects. The examinee is told that the examiner is thinking of one object, and that the examinee should ask the fewest possible number of yes/no questions in order to identify the object. The executive functions measured by this test are abstract thinking and the identification of categories.

The Tower Test requires the examinee to move a series of disks across three pegs in order to match a picture in the shortest amount of moves. The examinee is instructed to only move one disk at a time and to never place a large disk over a smaller one. The Tower Test measures spatial planning, rule learning, inhibition of impulsiveness, inhibition of perseverative responding, and maintaining cognitive set.

The Color-Word Interference Test consists of four conditions aimed at measuring inhibition and cognitive flexibility. Condition 1 requires participants to name color patches while under a time limit. Condition 2 involves quickly reading color words that are printed in black ink. Condition 3 requires participants to inhibit reading color words by reading a discordant ink color. Finally, Condition 4 involves switching back and forth between naming the color of the printed word and the dissonant colored ink.

The Verbal Fluency Test consists of the three conditions of Letter Fluency, Category Fluency, and Category Switching. In Condition 1, Letter Fluency, the examinee is asked to say words that begin with a particular letter as quickly as possible. In

Condition 2, Category Fluency, the participant is asked to verbalize words of a specific semantic category as quickly as possible. In Condition 3, Category Switching, the examinee is required to alternate saying words of two different categories as quickly as possible. This test is purported to measure the examinee's word fluency skills and cognitive flexibility.

The test of Design Fluency taps the functions of response inhibition, design fluency, and cognitive flexibility through the use of three conditions. Condition 1 requires the examinee to draw as many designs as possible in a series of boxes containing five dots. In Condition 2 the examinee is required to connect unfilled dots only and inhibit connecting any filled dots. In Condition 3 the examinee has to switch between drawing designs in filled and unfilled dots.

The Trail Making Test consists of five conditions which require the examinee to sequentially connect letters and numbers according to varying rules. Condition 1 assesses visual scanning and requires the examinee to quickly find and mark every number "three" on a page with multiple numbers and letters. Condition 2 requires the examinee to correctly sequence numbers while under a time limit. Condition 3 involves the ability to quickly sequence letters. Condition 4 measures cognitive flexibility, and requires the examinee to sequence numbers and letters, switching between a number and a letter each time. Condition 5 assesses simple motor speed without a sequencing or switching component.

For the current study, the following measures of executive functioning and/or attention are utilized: Word Context; Twenty Questions; Sorting; Tower; Color-Word

Interference, Conditions 3 & 4; Design Fluency, Conditions 2 & 3; Verbal Fluency, Condition 3; and Trail Making, Condition 4. For Twenty Questions, the Initial Abstraction score, which determines level of abstract thinking, is used in the current study. For Sorting, the Card Sorting Confirmed Correct Sort, which measures total confirmed correct sorts, was utilized. For all other tests, the total score for the selected condition/s was used.

Reliability. The D-KEFS technical manual provides split-half and test-retest reliability estimates for a majority of the subtests, separated by age groups. Use of split-half versus test-retest reliability coefficients depended on the type of task being measured. Overall, the split-half reliability estimates were extremely varied across the subtests and age groups. Split-half reliabilities were in the moderate to high range for the Verbal Fluency Test (.68-.90), the Sorting Test (.62-.81), the Color Word Interference Test (.62-.86), and the Twenty Questions Test – Initial Abstraction (.72-.87) (Homack et al., 2005).

Test-retest reliability coefficients were also reported using 101 examinees from the original normative sample, and had varied results (see Table 5). Practice effects were often observed across the different measures. Therefore, some of the test-retest reliabilities for the subtests are low, but many of the correlations are adequate (Delis et al., 2001).

Table 5

Delis-Kaplan Executive Function System, Median Test-Retest Reliability Coefficients

Test	Median r_{12}	Test (cont.)	Median r_{12}
Trail Making		Color Word Interference	
Condition 1	0.56	Condition 1	0.76
Condition 2	0.59	Condition 2	0.62
Condition 3	0.59	Condition 3	0.75
Condition 4	0.36	Condition 4	0.65
Condition 5	0.77	Sorting Test	
Verbal Fluency		Condition 1	0.51
Condition 1	0.80	Condition 2	0.60
Condition 2	0.79	Twenty Questions	0.43
Condition 3	0.52	Word Context	0.70
Design Fluency		Tower	0.44
Condition 1	0.58		
Condition 2	0.57		
Condition 3	0.32		

Validity. Evidence of validity was reported from numerous sources, including intercorrelations between individual D-KEFS tests, correlations with D-KEFS tests and established neurocognitive measures, and evidence from studies assessing the sensitivity of D-KEFS tests for use with clinical populations (Delis et al., 2001). The D-KEFS tests that were modified from established neurocognitive assessments have a long history of validity that is evidenced in numerous research studies over the past several decades.

Overall, intercorrelations between D-KEFS tests signified adequate convergent and discriminate validity due to expected positive and negative correlations (Homack et al., 2005). Evidence of validity in the D-KEFS was also tested and reported via correlations of the D-KEFS with the *California Verbal Learning Test- Second Edition* (CVLT-II; Delis et al., 2000) and the *Wisconsin Card Sorting Test* (WCST; Heaton et al., 1993), in addition to findings within clinical populations. Convergent validity was determined by correlations between the D-KEFS, the *California Card Sorting Test* (Delis, 1988), and the WCST, where moderate correlations were found. Evidence of discriminate validity was determined by a lack of correlation between the D-KEFS tests and the CVLT-II, which is a verbal memory test. Various validity studies with the D-KEFS also indicate that the individual measures of the D-KEFS are fairly accurate in distinguishing the clinical groups of fetal alcohol exposure, schizophrenia, chronic alcoholism, Alzheimer's disease, and Huntington's disease (Delis et al., 2001; Homack et al., 2005).

Review of Selected Tests

The WJ III COG, NEPSY, and D-KEFS were chosen for the current study because each contain identified measures of executive functioning that also report narrow executive functioning subcomponents associated with each measure. Subtests purported to measure areas of executive functioning have been selected for the current study. Table 6 lists each subtest that has been selected, as well as associated executive functioning tasks measured by each subtest.

Table 6

Tests and Associated Executive Functioning Areas Measured in Study

Test	Area of EF Measured	Test (cont.)	Area of EF Measured
WJ III COG		**D-KEFS**	
Concept Formation	Mental flexibility, problem solving	*Word Context*	Problem solving, mental flexibility
Planning	Planning, problem solving	*Twenty Questions*	Problem solving, concept formation, abstract reasoning
Pair Cancellation	Inhibition, sustained attention	*Sorting*	Mental flexibility, initiation, problem solving
Numbers Reversed	Attentional capacity, working memory	*Tower*	Planning, inhibition, establishing/ maintaining set, problem solving
Auditory Working Memory	Divided attention, working memory	*Color-Word Interference*	
Auditory Attention	Selective attention, inhibition	*Condition 3*	Inhibition
		Condition 4	Inhibition, mental flexibility, shifting attention
NEPSY			
Tower	Planning, self-regulation/ monitoring, inhibition, problem solving	*Design Fluency*	
		Condition 2	Inhibition
Auditory Attention	Selective attention, sustained attention, inhibition	*Condition 3*	Mental flexibility, shifting attention
Response Set	Mental flexibility, selective attention, sustained attention, inhibition	*Verbal Fluency*	
		Condition 3	Mental flexibility, shifting attention
Visual Attention	Sustained Attention	*Trail Making Test*	
		Condition 4	Mental flexibility, shifting attention

Data Analysis

Descriptive Statistics

Descriptive statistics were first calculated on each of the executive functioning subtests from the WJ III COG, NEPSY, and D-KEFS. Means, standard deviations, and ranges for these test scores were calculated using the Statistical Package for the Social Sciences (SPSS) 15.0 statistical program. In addition, issues of normality and linearity were determined and will be discussed in the subsequent chapter.

Bivariate Correlations

Bivariate correlations were also be conducted, using the SPSS 15.0 statistical program, in order to better understand the amount and degree of relationship between each subtest used in the study. Correlations were conducted between proposed subtests and correlation matrices including correlation coefficients and significance levels for the entire set of scores were developed. These analyses helped determine which tasks measure convergent abilities, evidenced by significant correlations, and which tasks measure discriminate abilities, evidenced by weak or insignificant correlations (McGrew & Woodcock, 2001).

Confirmatory Factor Analysis and Structural Equation Modeling

In order to determine how well various neurocognitive and executive functioning theories fit with executive functioning subtests of the WJ III COG, NEPSY, and D-KEFS, confirmatory factor analysis was utilized. Factor analysis is an approach to data analyses in which patterns of covariation among multiple observed variables are used to obtain data on latent variables (Byrne, 1989; Gorsuch, 1983). Two basic types of factor analysis are commonly employed: exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). EFA is used whenever the researcher does not already have knowledge regarding the latent variable structure of a group of data. Thus, with EFA, the researcher seeks to ascertain the minimal number of latent variables that explain the data. Conversely, CFA is employed whenever the researcher has a preexisting knowledge of the underlying latent variable structure that is based on theory, empirical research, or both theory and research (Byrne; Thompson, 2004). Therefore, CFA is often used to develop and validate psychological tests, where it is imperative to identify how well the test actually measures the underlying latent variables that it is hypothesized to measure (Anderson & Gerbing, 1988).

CFA is subsumed within a larger set of statistical techniques, known as structural equation modeling (SEM; Schumacker & Lomax, 2004). SEM employs multiple types of theoretical models in order to depict relationships among observed variables. SEM also establishes the degree to which theoretical models are supported by sample data (Schumacker & Lomax). Under the umbrella of SEM are several different types of modeling approaches, namely, regression models, path analysis models, and confirmatory factor models. Regression models allow for the prediction of dependent observed variable outcomes based on independent observed variable scores. Regression analysis assesses theoretical models in order to predict outcomes based on sample data (Schumacker & Lomax). Path models utilize correlation coefficients as well as regression analysis in order to model more complex relationships among observed variables. Path analysis depicts more meaningful theoretical relationships across observed variables than regression models (Schumacker & Lomax). Regression models and path models operate exclusively with observed variables, whereas confirmatory factor models include both observed and latent variables. In confirmatory factor modeling, a hypothesized model or set of models are tested to determine statistical significance, or in other words, if sample data confirms specified models. Confirmatory factor modeling was utilized in the current study.

There are three different approaches to SEM modeling that can be employed when determining how a theory will be analyzed (Schumacker & Lomax, 2004). In the confirmatory approach, one specific theoretical model is hypothesized and tested to determine whether the data fit the model. With this approach, the model is either rejected or accepted based on the outcome of the analysis. The next approach, the model generating approach, utilizes an initial, emerging model; however, the data generally does not fit the initial model at an acceptable level. Therefore, modification indices are used to add or delete paths in order to arrive at the best possible final model. The last approach, the alternative models approach, is the method that was utilized in the current study. In the alternative models approach, a limited number of theoretically differing models are selected in order to establish the model that best fits the sample data. In the current study, the selected models were tested within the same dataset; therefore, they can be referred to as nested models (Schumacker & Lomax).

There are five primary steps to SEM analysis that were followed in the current study, namely, model specification, model identification, model estimation, model testing, and model modification (Mueller, 1997; Schumacker & Lomax, 2004;

Thompson, 2000). The first step in any confirmatory factor modeling approach is to first specify models based on theoretical data and prior research. Researchers should utilize all available information to determine what variables and parameters should be included in the study (Anderson & Gerbing, 1988; Schumacker & Lomax). Prior to data collection, a researcher should specify a model to be confirmed with variance-covariance data. When building measurement models, it is important to structure the model so that each latent construct is defined by two or more indicators, or observed variables. It is also important that each observed measure is an estimate of only one latent construct. This allows for unidiminsional construct measurement with the most unambiguous inferences of meaning when determining latent construct outcomes (Anderson & Gerbing). For the current study, models based on available executive functioning research and theories were created according to these guidelines and can be found later in this chapter.

In model identification the goal is to determine if a unique set of parameter estimates are available, or in other words, if the data can fit only one theoretical model (Schumacker & Lomax, 2004; Thompson, 2004). This increases the meaningfulness of the data by providing confidence that the sample data does not fit an infinite number of theoretical models. In order to achieve model identification, in the current study, each parameter was specified depending on whether the parameter was known (fixed) or needed to be estimated (free). For example, a parameter would hypothetically be set as fixed if it is known that a specific observed variable does not load on a specific latent variable (e.g., an untimed subtest will not load on a processing speed construct). Therefore, a free parameter would be set if an observed variable is hypothesized to load on a latent variable, but the degree of the relationship still needs to be estimated. After each parameter was specified, the parameters were combined to form a model implied variance-covarience matrix (Schumaker & Lomax). The number of free parameters were then subtracted from the number of total possible parameters in the model, with the difference predicting the degrees of freedom. This aided in determining the level of model identification. A model can be considered underidentified if one or more parameters are not uniquely determined, due to limited information in the matrix (degrees of freedom < 0). A model may be just-identified if all parameters are uniquely determined due to just enough information in the matrix (degrees of freedom = 0). A model may be overidentified if there is more than one way to estimate the parameter due to more than enough information in the matrix (degrees of freedom > 0). If a model is overidentified or just-indentified the data analysis can proceed (Schumacker & Lomax).

The next step in SEM analysis is model estimation. In this step, estimates of the parameters in the implied variance-covariance matrix were selected, in order to determine if these parameter values yielded a matrix that is similar to the sample variance-covariance matrix of the observed variables (Mueller, 1997; Shumacker & Lomax, 2004). Once the parameters were estimated, model testing was completed to determine the fit of the implied matrix with the sample variance-covariance matrix. Specifically, a two-step modeling approach was utilized in the current study, which allows meaningful inferences to be made regarding theoretical constructs and their interrelations (Anderson & Gerbing, 1988). Within a two-step modeling approach the analysis of a measurement model is first tested, followed by the analysis of a structural model. The measurement model determines the relationships among the observed variables underlying the latent constructs. This allows for information to be obtained on convergent and discriminate validity evidence. The structural model indicates relationships among the latent constructs, as hypothesized by theory. The separate analysis of the structural model allows for information to be obtained regarding nomological (i.e., how well data truly represents constructs) and predictive validity (Anderson & Gerbing; Schumacker & Lomax).

In order to determine the significance of the similarities between the theoretical models and sample data, fit statistics were utilized. Fit statistics help to determine either the global fit of the implied matrix to the sample data or determine the fit of individual parameters (Schumacker & Lomax, 2004). The fit statistics that were utilized in the current study are described subsequently.

The chi-square (χ^2) test is a global statistic that determines the magnitude of discrepancy between the sample and the model covariance matrices. A χ^2 value that is not statistically significant indicates that the sample covariance matrix and the model's covariance matrix are similar, thus indicating adequate model fit. It is important to note that χ^2 tests are directly related to sample size and degrees of freedom. Therefore, the larger the sample size, the more likely a model will fail to fit sample data via χ^2 statistics (Barrett, 2007). A general rule of thumb in the literature is that whenever sample size exceeds 200, caution should be utilized whenever interpreting χ^2 test statistics (Barrett; Schumacker & Lomax, 2004). Due to this, an adjusted χ^2 test was utilized in the current study. The adjusted χ^2 test divides the χ^2 statistic by the degrees of freedom in the study, and helps account for large sample size (Wang, Fan, & Willson, 1996). Any adjusted χ^2 value less than 3.00 is considered acceptable. The adjusted χ^2 difference test was utilized to compare the fit of the alternative models in the current study.

Multiple fit indices are available to supplement the adjusted χ^2 test, each of which has unique strengths and weaknesses. Due to this, increased error is evident whenever fit indices are utilized singularly; therefore, it is important to analyze more than one fit index (Hu & Bentler, 1999). The fit indices that were utilized in the current study included the standardized root mean squared residual (SRMR), the root mean square error of approximation (RMSEA), the Non-Normed Fit Index (NNFI), and the comparative fit index (CFI). All of these indices are considered to be good indicators when using samples with non-random missing data. Furthermore, SRMR is considered to be the most sensitive index for models containing misspecified factor covariance or latent construct structures. RMSEA and CFI are considered to be sensitive indices to misspecificed factor loadings. Therefore, it is suggested that SRMR be used in conjunction with the other fit indices in order to obtain the best results with the least amount of error (Hu & Bentler). NNFI and RMSEA are considered beneficial because they take degrees of freedom into account (Bentler, 1990; Hu & Bentler). More specifically, RMSEA is beneficial when the number of variables increase and NNFI is preferred when testing alternate models (Shumacker & Lomax, 2004). Furthermore, CFI takes sample size into account, and is therefore more useful for studies with large sample sizes (Thompson, 2000).

When using fit indices, it is important to utilize predetermined cutoff criteria that aid in determining significant values. Adequate cutoff criteria for a fit index should result in limited Type 1 (i.e., rejecting the null hypothesis when it is true) and Type 2 (i.e., accepting the null hypothesis when it is false) error rates (Hu & Bentler, 1999). For SRMR and RMSEA, cutoff values less than .09 are considered acceptable. For NNFI and CFI, cutoff values greater than .70 are acceptable (Browne & Cudeck, 1993; Hu & Bentler; Yadama & Pandey, 1995).

In addition, individual parameter estimates for the paths in each model were also tested for statistical significance and reported as *t* scores. The magnitude and direction of the parameter estimates were then analyzed in order to determine if each significant path is theoretically meaningful, based on prior research (Schumacker & Lomax, 2004). For example, it would not be theoretically meaningful to have a negative coefficient between two constructs that are theoretically commensurate.

The last step in SEM analysis is model modification. Within this step, model specification is further taken into account in order to determine if a model should be altered in any way (Schumaker & Lomax, 2004; Thompson, 2000). A model is considered properly specified when the sample variance-covariation matrix is adequately reproduced by data from the implied theoretical model. The goal in model specification is to determine the theoretical model that most closely fits with the variation-covariation matrix from the sample data. If the sample variation-covariation matrix is not consistent with the data from the theoretical model, then the theoretical model is considered misspecified. Misspecification may be due to a significant variable or relationship that was omitted, or a nonsignificant variable or relationship that was included in the model. In order to determine the specification of the model, specification searches were completed first on each measurement model and then on the structural models (Anderson & Gerbing, 1988). The specification search process aids in determining if statistically significant parameters are practically significant and meaningful. Generally, when parameters of statistically significant models are not practically significant or meaningful, they are altered or deleted. One model in the current study was modified. After the modification, it was again tested for fit with the sample data. Any change in fit between models was then analyzed to determine which, if any models, significantly fit the data better than the others (Anderson & Gerbing; Schumacker & Lomax).

Furthermore, a single-sample cross-validation index (ECVI) was calculated for each of the models. ECVI is typically used whenever alternative models are analyzed within one set of data (Browne & Cudeck, 1993). The alternative model with the smallest ECVI value should prove to be the most stable in the population. The reliability of latent factors will also be calculated using *Cronbach Alpha* reliability coefficients (Schumacker & Lomax, 2004).

Therefore, for the current study, the analyses were completed by first entering demographic data and test scores into the SPSS 15.0 program. Next, the data was analyzed within the Lisrel 8.80 computer program, which utilizes visual representations of hypothesized models to outline the theoretical relationship between observed and latent variables. The Lisrel 8.80 program then determines the goodness-of-fit between the hypothesized model and the actual sample data (Byrne, 1989). By employing this technique, the current study compared several different executive functioning theoretical models in order to determine the model that best fit with the data from executive functioning tasks across the three neurocognitive test batteries. The research question and analyzed models are discussed subsequently.

Research Question

Is the underlying factor structure of executive functioning tasks across the WJ III COG, NEPSY, and D-KEFS in a clinical population of children best described by:

a. A theoretical model in which all tasks load on one, general executive functioning factor?
b. Anderson and colleagues' (2002) theory of executive functioning, with subtests loading on the three factors of attentional control, cognitive flexibility, and goal setting?
c. The CHC model of Cognitive Abilities (McGrew, 2005)?
d. The School Neuropsychological conceptual model (SNP model; Miller, 2007, 2010)?

A joint confirmatory factor analyses was conducted on the executive functioning tasks across the WJ III COG, NEPSY, and D-KEFS. This type of analysis combines and analyzes subtests of multiple measures using confirmatory factor analytic procedures (Schumacker & Lomax, 2004). The proposed models were analyzed separately and then compared to determine the model that best explains underlying executive functioning constructs. Evaluation of the level of fit for each model was determined by utilizing fit statistics. The following models were analyzed and compared in order to answer the aforementioned research question.

Model one (one general factor). For this model, all of the executive functioning tasks across the WJ III COG, NEPSY, and D-KEFS were forced to load on one, general, executive functioning factor. This simple, one-factor model serves as a comparison tool in contrast to more advanced models. It is desirable to rule out one-factor models so that support for multi-factor models is stronger (Thompson, 2004). Model one is presented in Figure 2.

Model two (Anderson and colleagues' model). Anderson and colleagues (2002) operationalized executive functions to factor into the components of attentional control, goal setting, and cognitive flexibility. For the current study, loadings of each subtest onto one of the three areas outlined by Anderson and colleagues have been determined based on the descriptions of each subtest from test authors (Delis et al., 2001; Korkman et al., 1998; Woodcock et al., 2001c), as well as through task analyses for each subtest.

Within this model, the following subtests were designated to load on the Attentional Control factor: WJ III COG Auditory Attention and Pair Cancelation, NEPSY: Auditory Attention, Response Set, and Visual Attention, and D-KEFS: Design Fluency, Condition 2 and Color Word Interference, Condition 3. The following subtests loaded onto the Cognitive Flexibility factor: WJ III COG: Concept Formation, Numbers Reversed and Auditory Working Memory and D-KEFS: Design Fluency, Condition 3; Verbal Fluency, Condition 3; Trail Making Test, Condition 4; and Color-Word Interference, Condition 4. The subtests that loaded on the Goal Setting factor are as follows, WJ III COG: Planning, NEPSY: Tower, and D-KEFS: Tower, Sorting Test, Word Context, and Twenty Questions. Anderson and colleagues' model and the designated subtests are presented in Figure 3.

Model three (CHC model). The Cattell-Horn-Carroll (CHC; McGrew, 2005) theory asserts that a three-tier hierarchical model of cognitive abilities can explain cognitive functioning. Within CHC theory, the second tier includes broad abilities that represent general cognitive abilities. The seven broad cognitive abilities included in the current study are Comprehension-Knowledge (*Gc*), Long-Term Retrieval (*Glr*), Visual-Spatial Thinking (*Gv*), Auditory Processing (*Ga*), Fluid Reasoning (*Gf*), Processing Speed (*Gs*), and Short-Term Memory (*Gsm*). These are the seven broad abilities that are most often associated with general neurocognitive abilities. Determination of the loading of each subtest onto one of these seven areas has been based on descriptions of each subtest provided by test authors (Delis et al., 2001; Korkman et al., 1998; Woodcock et al. 2001c). The loading of each subtest onto the CHC broad abilities is further supported by the neurocognitive demand task analyses presented by Flanagan and colleagues (2010).

Specifically, the WJ III COG Concept Formation test and the D-KEFS Sorting task were designated to load on the *Gf* factor; the WJ III COG Planning test, the NEPSY Tower and Visual Attention tests and the D-KEFS Tower, Design Fluency, Condition 2, and Design Fluency, Condition 3 subtests loaded on the *Gv* factor; The WJ III COG Pair Cancelation and the D-KEFS Trail Making Test, Condition 4 subtests loaded on the *Gs* factor; the WJ III COG Numbers Reversed and Auditory Working Memory tests and the

NEPSY Auditory Attention and Response Set subtests loaded on the *Gsm* factor; the D-KEFS Word Context and Twenty Questions tests loaded on the *Gc* factor; and the D-KEFS Color-Word Interference, Condition 3; Color-Word Interference, Condition 4; and the Verbal Fluency, Condition 3 tests loaded on the *Glr* factor.

As only one subtest loads on the *Ga* factor (WJ III COG Auditory Attention), that construct was tested as an independent variable. The model was tested both with and without the *Ga* factor variable, in order to determine which model represents the best fit with data. The CHC theory and designated subtests, with the *Ga* factor, are presented in Figure 4.

Model four (SNP model). The Conceptual Model for School Neuropsychological Assessment (SNP model; Miller, 2007, 2010) is a broad neurocognitive model that is specifically geared towards the assessment of neuropsychological processes in children. The SNP model suggests that the areas of executive functioning that can be measured by current neuropsychological assessments are as follows; Concept Generation, Inhibition, Motor Programming, Planning/ Reasoning/ Problem Solving, Set Shifting, Retrieval Fluency, Selective/ Focused Attention, Sustained Attention, the Use of Feedback in Task Performance, and Working Memory.

No measures of Motor Programming are included in the current study; therefore, this area was not included. Additionally, as outlined by the SNP model, the areas of Concept Generation and Retrieval fluency each include only one subtest from the current study; therefore, those areas were also not included in the current study. Furthermore, subtests identified under the Use of Feedback in Task Performance category can be better subtests identified under the Use of Feedback in Task Performance category can be better accounted for in other areas; thus, this category was also not included in the current study.

Therefore, the selected SNP model included tasks that load on the six categories of Selective/ Focused Attention, Sustained Attention, Inhibition, Set-Shifting, Planning/ Reasoning/ Problem Solving, and Working Memory. The loadings of each subtest onto one of these six areas have been based on the SNP model (Miller, 2007, 2010). Within the SNP model, some subtests are listed under multiple categories. For these subtests, task analysis further determined under which category they best fit.

Specifically, the WJ III COG Auditory Attention and Pair Cancellation subtests were designated to load on the Selective/ Focused Attention factor; the NEPSY Auditory Attention and Visual Attention tests loaded on the Sustained Attention factor; the D-KEFS Color-Word Interference, Condition 3 and NEPSY Response Set subtests will loaded on the Inhibition factor; the D-KEFS Color-Word Interference, Condition 4; Design Fluency, Condition 3; Verbal Fluency, Condition 3; and Trail Making Test, Condition 4 subtests loaded on the set-shifting factor; the WJ III COG Concept Formation and Planning tests, the NEPSY Tower test, and the D-KEFS Twenty Questions, Tower, and Word Context subtests loaded on the Planning/ Reasoning/ and Problem Solving Factor, and the WJ III COG Numbers Reversed and Auditory Working Memory subtests loaded on the Working Memory Factor. The SNP model and designated subtests are presented in Figure 5.

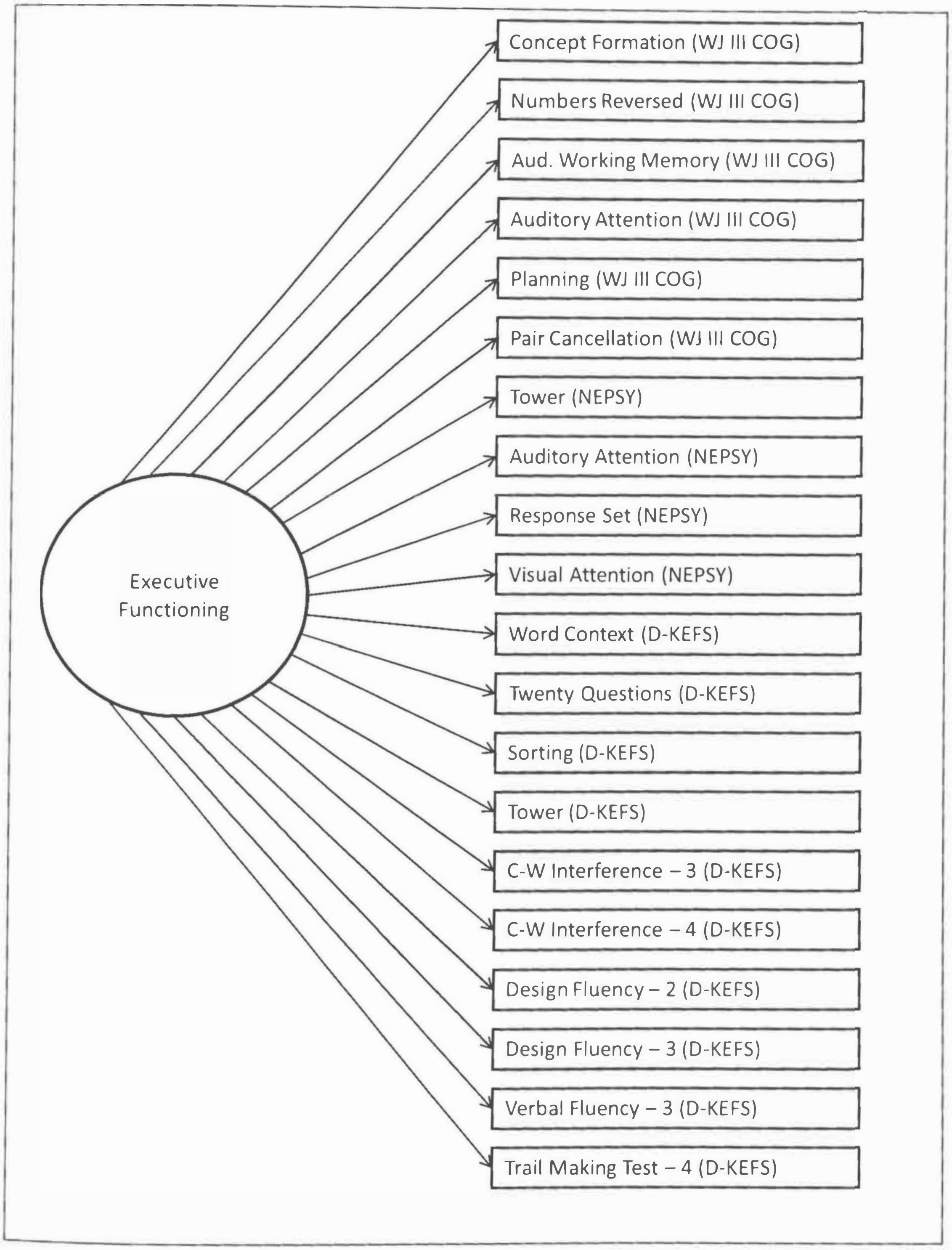

Figure 2. Illustration of the general factor model (Model 1)

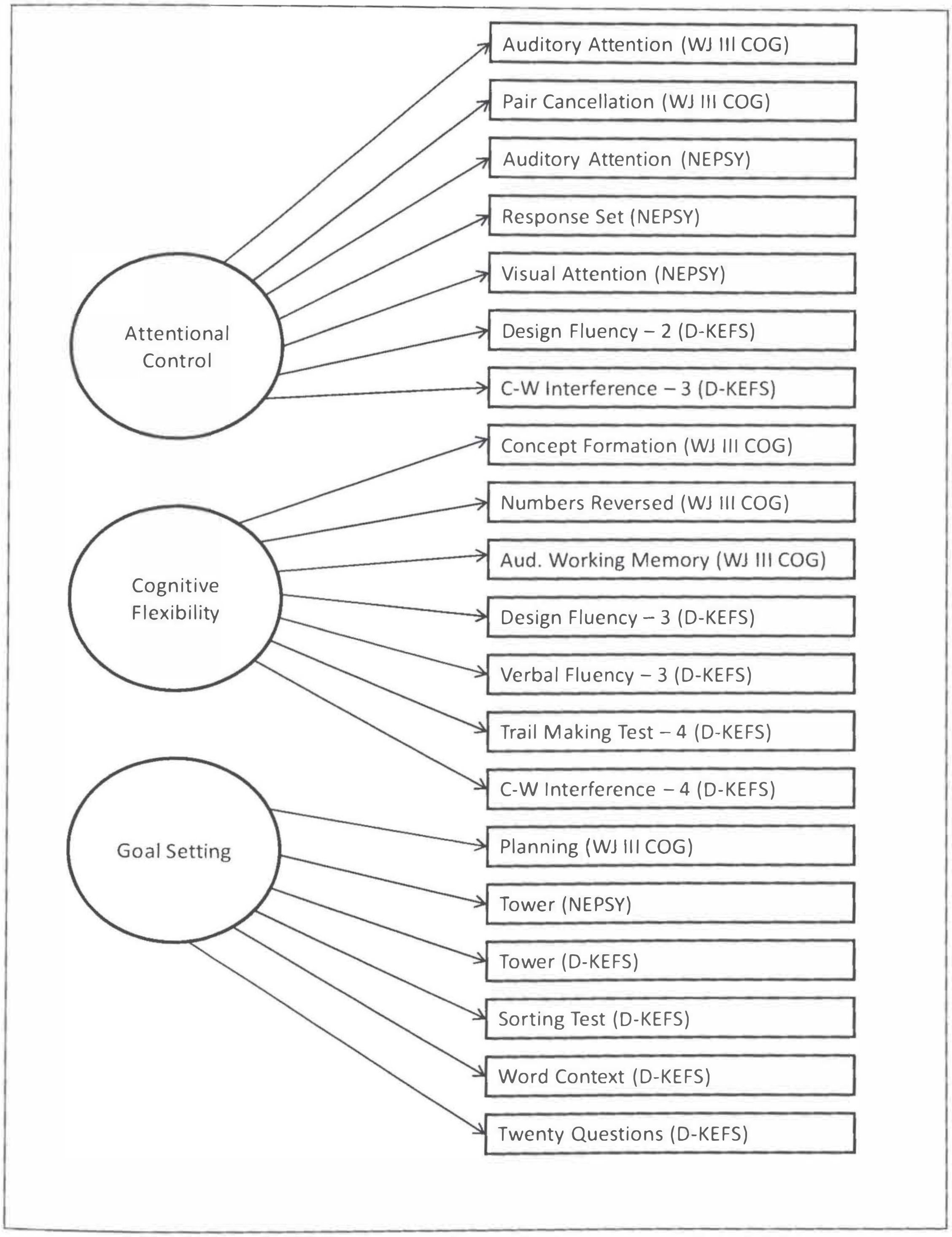

Figure 3. Illustration of the Anderson et al. (2002) factor model (Model 2)

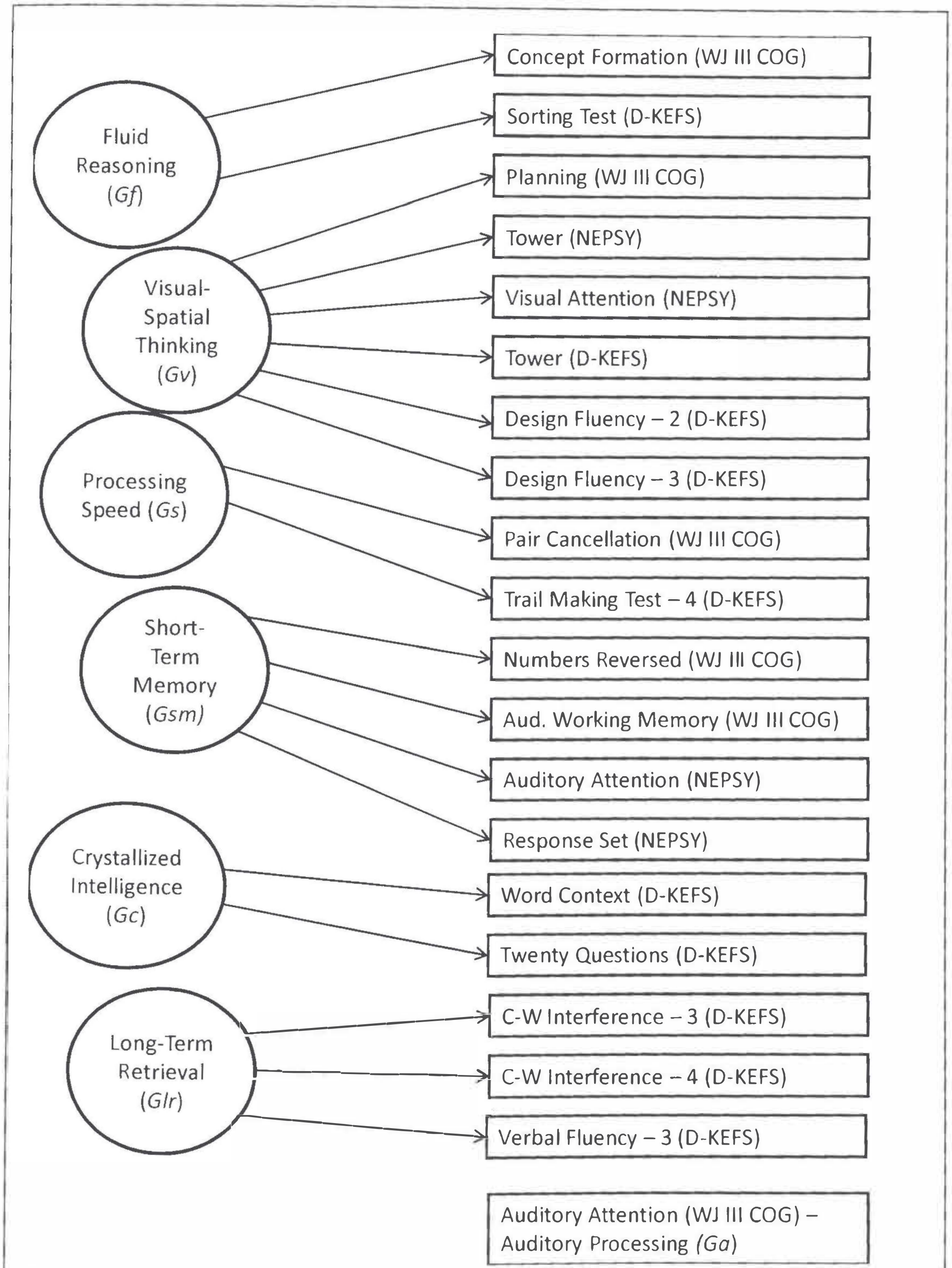

Figure 4. Illustration of the CHC theory factor model (Model 3)

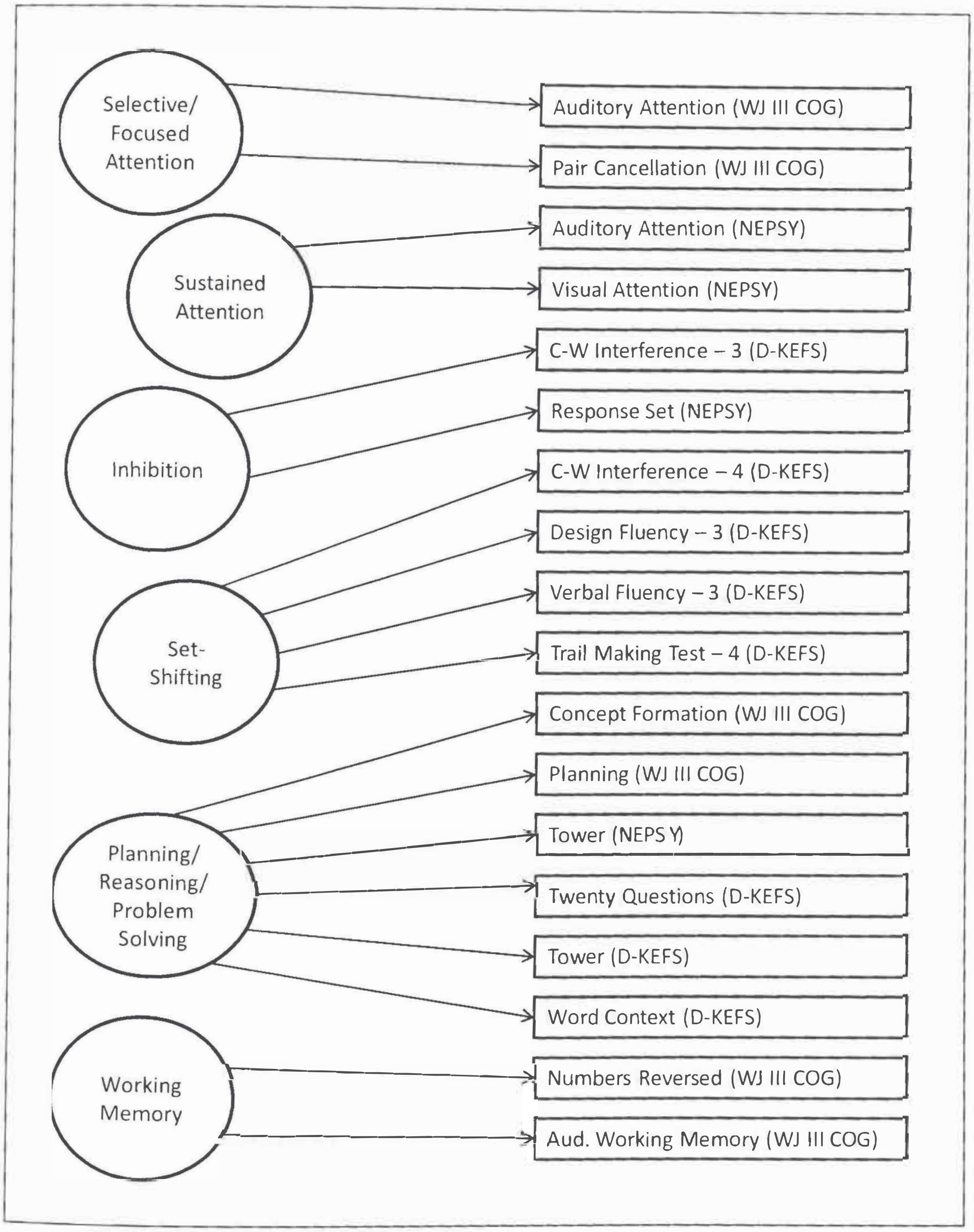

Figure 5. Illustration of the School Neuropsychology factor model (Model 4)

Methodological Issues

The current study utilized archival data that was collected from multiple professionals. Additionally, data collected was from individualized school neuropsychological evaluations administered to a wide range of children. Due to this, the occurrence of missing data was expected in the study, which could directly affect statistical analysis and outcomes. Specifically, with missing data, statistical power can be reduced due to the loss of data. Also, missing data can produce biased parameter estimates (Allison, 2003). While using only the complete cases (i.e., listwise deletion) is the simplest solution, information is lost by not using the incomplete cases. This is particularly the case since missing data in the current sample was not be missing at random (MNAR). Furthermore, when incomplete cases are deleted through listwise deletion, any methodical differences between the complete cases and the incomplete cases may fail to be detected. Accordingly, the resulting inferences may not be applicable to the population of all cases, especially with a smaller number of complete cases (Allison; Schafer & Olsen, 1998).

Therefore, multiple imputation (MI) was utilized to account for missing data. Within this method each missing datum is replaced by $m>1$ plausible values drawn from the predictive distribution of missing data under the appropriate data model and the missing data mechanism. The result is m completed data sets, which were analyzed separately and then combined using Rubin's (1987) rule to produce one overall inference that accounted for the missing data uncertainty. The primary advantage of the MI method is that it leads to valid statistical inference in the presence of non-response. In addition, MI produces approximately unbiased parameter estimates under reasonable assumptions. To implement MI for the present data set, each missing value was imputed five times using Lisrel 8.80.

The cases used in the current study were also comprised of a variety of neuropsychological clinical groups, thereby creating a sample that was not representative of the population. Consequently, the results of this study may not be generalizable to other populations or circumstances. However, it should also be noted that because each of the professionals that contributed data were trained through the same program, consistency among testing procedures and techniques was expected. Furthermore, the use of archival data allows for a broad collection of participants from varied geographic regions and demographic backgrounds, which may positively impact the generalizability of the study.

Although limited sample size was not expected to be problematic for the current study, sample size is a significant methodological issue that still needs to be discussed. Limits in sample size may have affected the accuracy of estimations and fit statistics of the current study (Schumacker & Lomax, 2004). Although multiple numeric rules of thumb for minimum sample size have been reported in the literature, the minimum sample size for adequate power depends on more than a cutoff value alone (MacCallum, Browne, & Sugawara, 1996; MacCallum, Widaman, Zhang, & Hong, 1999). For example, the ratio of sample size to the number of factors and variables has also been demonstrated as being important. Comrey and Lee (1992) stated that the sample size should be five times as large as the numbers of variables and factors in order to achieve adequate power. Nunnally (1967) indicates that when using confirmatory factor analysis, at least ten cases per factor are needed for adequate power. Other factors, such as the degree of communality (i.e., the proportion of variance each factor has in common with other factors) and level of identification have also been suggested as important variables in determining adequate sample size (MacCallum et al., 1999). According to these standards, the sample size of 321 utilized in the current study indicates adequate statistical power. However, as already discussed, since the sample size in the current study was above 300, methodological issues with various fit indices may surface. Therefore, a variety of fit indices, some of which are not affected by sample size, were utilized in the current study.

Chapter Summary

The current study was designed to assess the level of concurrent validity of executive functioning tasks among commonly used neurocognitive assessment batteries. This study was also created in order to determine the applicability and fit of existing executive functioning theories to a clinical population of subjects. The preceding review of the literature provided confirmation for a need to research the validity of executive functioning components of the WJ III COG, NESPY, and D-KEFS, as well as their fit with theories of executive functioning. This was examined by executing a series of confirmatory factor analyses in order to determine the underlying constructs of executive functioning tasks across the WJ III COG, NEPSY, and D-KEFS.

Subtests scores were obtained through archival data collected from case studies submitted as part of course completion requirements for the KIDS, Inc.'s School

Neuropsychology Post-Graduate Certification Program. Due to the selected assessment instruments, the current study utilized information from a mixed clinical sample of children between the ages of eight and twelve. Descriptive statistics were first calculated on the designated subtest scores of the WJ III COG, NEPSY, and D-KEFS and bivariate correlations were conducted in order to better understand the amount and degree of relationship between each subset used in the current study.

The study also utilized confirmatory factor analysis in order to determine the underlying executive functioning constructs across the three test batteries. Specifically, four models were analyzed and compared to determine the best fit with the sample data. These models include a general factor model, a factor model arising from Anderson and colleagues' (2002) theory of executive functioning, a model developed from CHC theory (McGrew, 2005), and a model outlining the SNP model (Miller, 2007, 2010). All of these models were analyzed using confirmatory factor analytic procedures from the Lisrel 8.80 statistical program.

CHAPTER IV

RESULTS

The aim of this study was to examine the underlying constructs of three neurocognitive tests in relation to three theories of executive functioning. The factor structures of the *Woodcock Johnson III Tests of Cognitive* Abilities (*WJ III COG*; Woodcock et al., 2001c), *The NEPSY, A Developmental Neuropsychological Assessment* (NEPSY; Korkman, Kirk, & Kemp, 1998), and the *Delis-Kaplan Executive Function System* (*D-KEFS*; Delis et al., 2001) were evaluated through a joint confirmatory factor analysis. Anderson, Levin, and Jacob's (2002) theory of executive functioning, the Cattell-Horn-Carroll theory of cognitive abilities (CHC theory; McGrew, 2005; McGrew & Woodcock, 2001), the School Neuropsychological (SNP) conceptual model (Miller, 2007, 2010), and a one-factor executive functioning model, used as a baseline comparison, were analyzed separately and then compared to determine the model that best explains underlying executive functioning constructs.

Preliminary Analysis

Descriptive Statistics

Before conducting the statistical analysis, missing data was identified, and as it was not missing at random, multiple imputation was utilized to account for any missing values (Rubin, 1987). The data was also analyzed for linearity and kurtosis. The distributions of all subscales followed a relatively normal shape and were not skewed to the left or the right. Kurtosis, or a measure of the peakedness or flatness of each distribution was also assessed. Values closer to zero represent a normal distribution. A kurtosis value of +/-1 is considered very good, but +/-2 is also considered acceptable for most psychometric uses (Joanes & Gill, 1998). All subtests utilized in the current sample reflected very good to adequate distributions. Three subtests (e.g., D-KEFS Verbal Fluency, Condition 2; Verbal Fluency, Condition 3; and Design Fluency, Condition 3) reflected slight positive skewness (1.73, 1.47, and 1.23, respectively), suggesting that these subtests include a greater number of outlier scores at each end of the distribution. All other subtests were within the +/-1 range.

Frequencies and percentages were also calculated for four categorical demographic variables. Information regarding gender, language, and ethnicity is reported in Table 7. The majority of the population used in the study was comprised of males (68.3%) whereas females made up 31.1% of the population. Gender was not reported for two cases. Given the use of a mixed, clinical sample of participants, and research demonstrating the higher frequency of many school-aged clinical disorders in males (Gershon, 2002), the discrepancy among gender in this sample is as expected. Furthermore, the majority of subjects used in the study were primarily English speaking (67.9%), followed by primarily Spanish speaking (5.1%). Primary language was not reported in 26.9% of the cases. Although 65.4% of the sample did not have a reported ethnicity; Caucasian students (20.5%) made up the largest percentage of the sample when ethnicity was reported. This was followed by Hispanic/ Latino American (8.3%), African-American (3.2%), Bi-Racial (1.6%), Foreign National (0.6%), and Asian American (0.3%) students.

Table 7

Frequencies and Percentages for Child's Gender, Language, and Ethnicity

	N	%
Child's Gender		
Female	97	31.1
Male	213	68.3
Not Reported	2	0.6
Primary Language		
English	212	67.9
Spanish	16	5.1
Not Reported	84	26.9
Ethnicity		
Caucasian	64	20.5
African-American	10	3.2
Hispanic/ Latino American	26	8.3
Asian-American	1	0.3
Bi-Racial	5	1.6
Foreign National	2	0.6
Not Reported/ Unknown	204	65.4

Table 8 outlines the frequencies and percentages based on child's diagnosis. The majority of the children utilized in the current student, who had listed clinical diagnoses, have been diagnosed with a learning disability (24.0%), followed by students with multiple disabilities (15.4%), Attention Deficit Hyperactivity Disorder (ADHD) or

Attention Deficit Disorder (ADD; 14.1%), neurological impairment (10.6%), and an Autism Spectrum Disorder (ASD; 4.6%). The category of neurological impairment was comprised of students with traumatic or acquired brain injuries, seizure disorders, brain tumors, or miscellaneous neurological impairments. Clinical diagnosis was not reported for 21.8% of the population. Refer to Table 8 for a list of less frequent clinical diagnoses, and combinations of diagnoses, utilized in the study.

Table 8

Frequencies and Percentages for Child's Diagnosis

	N	%
Child's Diagnosis		
Learning Disability	75	24.0
Language Disability	11	3.5
Mental Retardation	2	0.6
Neurological Impairment	33	10.6
ADHD/ ADD	44	14.1
Autism Spectrum Disorder	15	4.8
Emotional Disability	8	2.6
General Medical	7	2.2
Deaf	1	0.3
Multiple Disabilities	48	15.4
Not Reported/ Unknown	68	21.8

Note. ADHD = Attention Deficit Hyperactivity Disorder; ADD = Attention Deficit Disorder.

The make-up of clinical diagnoses used in the study is not representative of frequencies of diagnoses in the general population. Rather, the diagnoses in the clinical sample are more neurologically-based in nature, which is as expected given the neuropsychological basis of the assessments administered to this sample. For other statistical analyses utilized in this study, clinical diagnosis was grouped based on the highest frequency of occurrence in this sample (specifically, learning disability, language disability, ADHD/ ADD, ASD, multiple disabilities, or unknown/ unreported disabilities). Neurological impairment was not specified in further analyses due to its inclusion of multiple subgroups (i.e., traumatic or acquired brain injuries, seizure disorders, brain tumors, etc.), that when separated, resulted in low *N*'s.

Frequencies, means, standard deviations (*SD*), and ranges for each subtest utilized in the analysis were calculated for the sample and are presented in Table 9. The standardized mean for the WJ III COG Standard Score subtests is 100, with a *SD* of 15. The standardized mean of the NEPSY and D-KEFS Scaled Score subtests is 10, with a *SD* of 3. Means are described as falling within the average range if they fall between the 25^{th} and 74^{th} percentile (between 90 to 109 for Standard Scores and between 8 to12 for Scaled Scores). The means of each of the analyzed WJ III COG subtests fell within one standard deviation of the standardized mean, with means ranging from 90.22 to 100.67. The means of three of the NEPSY subtests fell within the average range, with the exception of Response Set, which fell below the average range. Due to the mixed clinical sample of children utilized in this study, this deviation is expected, as the NEPSY Response Set subtest requires higher tasks demands that may prove more difficult for children with clinical diagnoses.

Table 9

Means and Standard Deviations for Continuous Subtest Variables

	N	Mean	*SD*	Min	Max
WJ III COG					
Auditory Attention	308	97.00	10.43	69	124
Pair Cancellation	309	94.23	9.67	71	120
Auditory Working Memory	312	96.21	13.17	61	131
Numbers Reversed	312	90.22	12.15	56	122
Concept Formation	312	95.27	14.41	59	138
Planning	300	100.67	7.12	82	120
NEPSY					
Auditory Attention	311	9.15	2.32	2	14
Response Set	312	7.91	2.63	1	13
Visual Attention	312	9.09	2.96	1	19
Tower	310	9.08	2.70	3	16
D-KEFS					
Color Word Interference 3	312	7.20	2.93	1	13
Trail Making Test 4	312	7.06	2.55	1	15
Verbal Fluency 3	301	8.13	1.97	2	13
Color Word Interference 4	309	7.95	2.21	3	13
Design Fluency 3	312	8.11	2.04	2	14
Card Sorting	312	8.42	2.95	2	16
Twenty Questions	302	8.64	2.16	4	15
Tower	311	9.43	2.52	3	15
Word Context	303	7.31	1.96	3	12
Design Fluency 2	312	9.38	2.02	5	15

The means of six of the D-KEFS subtests fell within the average range, with the means of four subtests falling below the average range. Some of the subtests that required inhibition and mental flexibility (Color Word Interference, Condition 3; Color Word

Interference, Condition 4; Trail Making Test, Condition 4; and Word Context) resulted in lower overall scores in children with clinical diagnoses. For most subtests, the sample was restricted in range (*SD* of the WJ III COG subtest scores = 7.12 to 14.41; *SD* of NEPSY subtest scores = 2.32 to 2.96; *SD* of D-KEFS test scores = 1.96 to 2.95), which was unexpected given the mixed, clinical sample of children. Since children with a variety of clinical diagnosis were included in the sample, more variability would be expected in test scores. The bivariate correlation coefficients between subtests, described later in the chapter, were likely higher due to this restriction in range.

Relationships among Demographic Variables

Crosstabulation analysis using Pearson's chi-square and Cramer's V were conducted to examine the relationship between participants' gender, language, ethnicity and broad diagnosis. As shown in Table 10, there were no significant relationships for gender with language, ethnicity, or broad diagnosis, all p's *ns*. Table 11 shows the relationship for language with ethnicity and broad diagnosis. A significant relationship was found between ethnicity and language, $\chi^2(6) = 43.76$, $p = <.001$, Cramer's $V = .27$. A greater proportion of those who spoke English were Caucasian (25.1%) compared to those who did not speak English (5.3%) or whose language was not reported (14.3%). In contrast, a greater proportion of those who did not speak English were Hispanic/Latino American (47.4%) compared to those who spoke English (5.8%) or whose language was not reported (6.5%).

Table 10

Frequencies and Percentages for Language, Ethnicity, and Broad Diagnosis by Gender

	Female		Male			
	n	%	n	%	χ^2	*p*
Language					1.65	.438
English	71	73.2	141	66.2		
Non-English	6	6.2	14	6.6		
Not Reported	20	20.6	58	27.2		
Ethnicity					2.78	.428
Caucasian	23	24.5	41	19.8		
African-American	1	1.1	9	4.3		
Hispanic/LA	8	8.5	18	8.7		
Not Reported	62	66.0	139	67.1		
Broad Diagnosis					5.73	.333
Learning Disability	25	28.4	49	23.9		
Language Disability	3	3.4	8	3.9		
ADHD/ ADD	14	15.9	19	9.3		
ASD	12	13.6	32	15.6		
Multiple Disabilities	17	19.3	61	29.8		
Not Reported	17	19.3	36	17.6		

Note. LA = Latino American, ADHD = Attention Deficit Hyperactivity Disorder; ADD = Attention Deficit Disorder, ASD = Autism Spectrum Disorders.

In addition, a significant relationship was found between broad diagnosis and language, $\chi^2(10) = 37.64$, $p = <.001$, Cramer's $V = .25$. A greater proportion of those who spoke English had a learning disability diagnosis (27.9%) compared to those who did not report a language (21.6%) or did not speak English (15.8%). Similarly, a greater proportion of those with an ASD spoke English (15.4%) or were without a reported language (17.6%), compared to those who did not speak English (0.0%). A greater proportion of those with multiple disorders spoke English (26.9%) or did not report a language (28.4%) compared to those who did not speak English (15.8%). In contrast, a greater proportion of those who spoke non-English had a language disorder diagnosis (26.3%) compared to those who spoke English (2.5%) or did not report a language (1.4%).Table 12 shows the relationships of broad diagnosis with ethnicity. No significant relationships were found for broad diagnosis with ethnicity, all *p*'s *ns*.

Table 11

Frequencies and Percentages for Ethnicity and Broad Diagnosis by Primary Language

	English		Non-English		Not Reported			
	n	%	n	%	n	%	χ^2	*p*
Ethnicity							43.8	<.001
Caucasian	52	25.1	1	5.3	11	14.3		
African-American	7	3.4	0	0.0	3	3.9		
Hispanic/ LA	12	5.8	9	47.4	5	6.5		
Not Reported	136	65.7	9	47.4	58	75.3		
Broad Diagnosis							37.6	<.001
Learning Disability	56	27.9	3	15.8	16	21.6		
Language Disability	5	2.5	5	26.3	1	1.4		
ADHD/ ADD	18	9.0	3	15.8	12	16.2		
ASD	31	15.4	0	0.0	13	17.6		
Multiple Disabilities	54	26.9	3	15.8	21	28.4		
Not Reported	37	18.4	5	26.3	11	14.9		

Note. LA = Latino American, ADHD = Attention Deficit Hyperactivity Disorder; ADD = Attention Deficit Disorder, ASD = Autism Spectrum Disorders.

Table 12

Frequencies and Percentages for Broad Diagnosis by Ethnicity

	Caucasian		African-American		Hispanic/LA		Not Reported			
	n	%	n	%	n	%	n	%	χ^2	*p*
Broad Diagnosis									16.07	.378
Learning Disability	20	33.9	0	0.0	6	24.0	48	24.9		
Language Disability	0	0.0	0	0.0	1	0.0	9	0.0		
ADD/ADHD	2	0.0	1	12.5	4	16.0	24	12.4		
ASD	11	18.6	1	12.5	6	24.0	26	13.5		
Multiple Disabilities	14	23.7	4	50.0	5	20.0	51	26.4		
Not Reported	12	20.3	2	25.0	3	12.0	35	18.1		

Note. LA = Latino American, ADHD = Attention Deficit Hyperactivity Disorder; ADD = Attention Deficit Disorder, ASD = Autism Spectrum Disorders.

Bivariate Correlations

Pearson-product moment correlations were computed between and among all of the subtests scores from the WJ III COG, NEPSY, and D-KEFS that were examined in this study. Intra- and inter-correlation between the WJ III COG and NEPSY subtest scores are presented in Table 13. Within the WJ III COG subtest scores, the correlations ranged from r = .183 (Planning and Pair Cancellation) to r = .507 (Concept Formation and Auditory Working Memory). Of the 15 intercorrelations calculated among the WJ III COG subtests, all of them were significant at the .01 level.

Table 13

Pearson's Product Moment Correlations between Woodcock Johnson III: Tests of Cognitive Abilities (WJ III COG) and NEPSY Subscales

	WJ III COG						NEPSY		
	AA	PC	AWM	NR	CF	PL	AA	RS	VA
WJ III COG									
PC	.205 **								
AWM	.347 **	.243 **							
NR	.235 **	.408 **	.483 **						
CF	.245 **	.396 **	.507 **	.501 **					
PL	.323 **	.183 **	.243 **	.214 **	.297 **				
NEPSY									
AA	.011	.103	.181 **	.224 **	.127 *	-.011			
RS	.076	.216 **	.218 **	.236 **	.237 **	-.010	.539 **		
VA	-.004	.295 **	.097	.178 **	.113 *	.060	.083	.270 **	
TW	.169 **	.142 *	.260 **	.172 **	.195 **	.131 *	.189 **	.176 **	.174 **

Note. AA = Auditory Attention; PC = Pair Cancelation; AWM = Auditory Working Memory; NR = Numbers Reversed; CF = Concept Formation; PL = Planning; RS = Response Set; VA = Visual Attention; TW = Tower.
* $p < .05$. ** $p < .01$

Within the NEPSY subtests scores, the correlations ranged from *r* = .083 (Auditory Attention and Visual Attention) to *r* = .539 (Auditory Attention and Response Set). Out of six intercorrelations calculated among the NEPSY, five were significant at the .01 level. The correlation between the NEPSY Auditory Attention and Visual Attention subtests was not significant at the .01 or .05 level.

Correlations were also examined between the NEPSY and WJ III COG subtests. Subtest correlations between these two measures ranged from *r* = -.011 (NEPSY Auditory Attention and WJ II COG Planning) to *r* = .295 (NEPSY Visual Attention and WJ III COG Pair Cancellation). Out of the 24 measured correlations, 12 were significant at the .01 level, 4 were significant at the .05 level, and 8 were not significant at either level. The NEPSY Tower test demonstrated the highest frequency of significantly correlated (at the .01 level) WJ III COG subtests. The WJ III COG Numbers Reversed subtest demonstrated the highest frequency of significantly correlated (at the .01 level) NEPSY tests.

Correlations among the subtests of the D-KEFS are presented in Table 14. Within D-KEFS subtests scores, the correlations ranged from *r* = -.119 (Tower Test and Verbal Fluency Test, Condition 3) to *r* = .471 (Color Word Interference Test, Conditions 3 and 4). Out of 45 intercorrelations calculated among the D-KEFS, fourteen were significant at the .01 level and nine were significant at the .05 level. Twenty two of the D-KEFS intercorrelations were not significant at either level.

Table 14

Pearson's Product Moment Correlations within the Delis-Kaplan Executive Function System (D-KEFS) Subscales

	CWI-3	TMT-4	VFT-3	CWI-4	DFT-3	CS	20Q	TW	WC
TMT-4	.071								
VFT-3	-.051	.095							
CWI-4	.471 **	.192 **	.141 *						
DFT-3	.06	.154 **	.128 *	.099					
CS	.215 **	.166 **	.014	.264 **	.245 **				
20Q	.134 *	-.016	.107	.120 *	-.093	.360 **			
TW	-.015	.071	-.119 *	.044	.206 **	.134 *	.030		
WC	.102	.194 **	.005	.090	.020	.109	.171 **	.019	
DFT-2	-.048	.136 *	.039	.088	.454 **	.290 **	.135 *	.133 *	.189 **

Note. CWI-3 = Color Word Interference, Condition 3; TMT-4 = Trail Making Test, Condition 4; VFT-3 = Verbal Fluency Test, Condition 3; CWI-4 = Color Word Interference, Condition 4; DFT-3 = Design Fluency Test, Condition 3; CS = Card Sorting; 20Q = Twenty Questions; TW = Tower Test; WC = Word Context; DFT-2; Design Fluency Test, Condition 2.

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

The large number of insignificant intercorrelations among D-KEFS subtests may be explained by the non-theoretical foundation of the D-KEFS. The D-KEFS was not created based on any one, unified theory, and each of the D-KEFS tests were designed as stand-alone measures of executive functioning. Therefore, a high number of intercorrelations are not to be expected.

Correlations between the D-KEFS and the WJ III COG and NEPSY are reported in Table 15. Subtest correlations between the D-KEFS and the WJ III COG ranged from $r = -.177$ (D-KEFS Card Sorting and WJ III COG Auditory Attention) to $r = .264$ (D-KEFS Color Word Interference, Condition 4 and WJ III COG Pair Cancellation). Out of the 60 measured correlations, 16 were significant at the .01 level, 10 were significant at the .05 level, and 34 were not significant at either level. Out of the D-KEFS subtests, Design Fluency Test, Condition 2 had the highest frequency of significant correlations (at the .01 level) with WJ III COG tests, followed by Trail Making Test, Condition 4. Out of the WJ III COG subtests, Pair Cancellation and Concept Formation had highest frequency of significant correlations (at the .01 level) with D-KEFS subtests. Subtest correlations between the D-KEFS and the NEPSY ranged from $r = -.091$ (D-KEFS Tower and NEPSY Tower) to $r = .332$ (D-KEFS Color Word Interference, Condition 4 and NEPSY Response Set). Out of the 40 measured correlations, 6 were significant at the .01 level, 6 were significant at the .05 level, and 28 were not significant at either level. Out of the D-KEFS subtests, Color Word Interference Test, Condition 3 had the highest frequency of significant correlations with NEPSY tests. Out of the NEPSY subtests, Response Set had highest frequency of significant correlations with D-KEFS subtests.

Table 15

Pearson's Product Moment Correlations between Delis-Kaplan Executive Function System (D-KEFS) and Woodcock Johnson III: Tests of Cognitive Abilities (WJ III COG) and NEPSY subscales

	WJ III COG						NEPSY			
	AA	PC	AWM	NR	CF	PL	AA	RS	VA	TW
D-KEFS										
CWI-3	-.044	.161 **	.160 **	.134 *	.053	-.037	.142 *	.184 **	.225 **	.186 **
TMT-4	.093	.241 **	.132 *	.177 **	.184 **	.013	.115 *	.098	.134 *	.126 *
VFT-3	-.097	.029	.106	.039	.095	-.024	.051	.040	-.014	.010
CWI-4	-.034	.264 **	.123 *	.143 *	.198 **	-.078	.240 **	.332 **	.106	.031
DFT-3	-.083	.201 **	.043	.070	.059	.127 *	.070	.017	.084	.130 *
CS	-.177 **	.140 *	.116 *	.110	.112 *	-.069	.104	.181 **	.103	.102
20Q	-.010	.008	-.039	.010	.015	-.151 **	.005	.077	.039	-.053
TW	-.015	.133 *	.060	.071	.085	.019	.008	.011	.008	-.091
WC	.101	.100	.121 *	.167 **	.183 **	-.104	.109	.133 *	.074	.084
DFT-2	.064	.086	.165 **	.203 **	.164 **	.193 **	.024	.094	.092	.084

Note. AA = Auditory Attention; PC = Pair Cancelation; AWM = Auditory Working Memory; NR = Numbers Reversed; CF = Concept Formation; PL = Planning; RS = Response Set; VA = Visual Attention; TW = Tower; CWI-3 = Color Word Interference, Condition 3; TMT-4 = Trail Making Test, Condition 4; VFT-3 = Verbal Fluency Test, Condition 3; CWI-4 = Color Word Interference, Condition 4; DFT-3 = Design Fluency Test, Condition 3; CS = Card Sorting; 20Q = Twenty Questions; WC = Word Context; DFT-2; Design Fluency Test, Condition 2.
* p < .05. ** p < .01

Relationships between the Measure Subscales and the Demographic Variables

In order to determine the multivariate differences between gender and the varying subscales of the WJ III COG, NEPSY and D-KEFS, a series of Multivariate Analysis of Variance (MANOVAs) were conducted (see Table 16). For the WJ III COG, there was a marginally significant multivariate effect between the two sexes, $F(6, 285) = 1.80$, $p = .099$. Further analysis showed significant differences on both the Pair Cancellation subscale, $F(1, 290) = 5.47$, $p = .021$, and the Concept Formation subscale, $F(1, 290) = 4.07$, $p = .045$. In terms of the Pair Cancellation subscale, females scored significantly higher ($M = 95.74$, $SD = 9.60$) than males ($M = 92.98$, $SD = 9.27$). Similarly, females scored significantly higher on the Concept Formation subscale ($M = 97.72$, $SD = 14.91$) than males ($M = 94.14$, $SD = 13.72$). No other significant differences were found on the other subscales, *ns*.

For the NEPSY, there was a significant multivariate effect between the two sexes, $F(4, 302) = 5.78$, $p < .001$. Further analysis showed significant differences on the Auditory Attention subscale, $F(1, 305) = 7.86$, $p = .005$, the Response Set subscale, $F(1, 305) = 18.21$, $p < .001$, and the Visual Attention subscale, $F(1, 305) = 8.67$, $p = .003$. In terms of the Auditory Attention subscale, females scored significantly higher ($M = 9.66$, $SD = 2.27$) than males ($M = 8.87$, $SD = 2.30$). Similarly, females scored significantly higher on the Response Set subscale ($M = 8.81$, $SD = 2.61$) than males ($M = 7.48$, $SD = 2.50$). Females also scored significantly higher on the Visual Attention subscale ($M = 9.75$, $SD = 3.08$) than males ($M = 8.73$, $SD = 2.67$). No significant difference was found on the Tower subscale, *ns*.

Table 16

Means and Standard Deviations for Woodcock Johnson III: Tests of Cognitive Abilities (WJ III COG), NEPSY, and Delis-Kaplan Executive Function System (D-KEFS) Subtest Scores by Child's Gender

	Female		Male			
	Mean	*SD*	Mean	*SD*	*F*	*P*
WJ III COG					1.80 *	.099
AA	96.34	11.25	97.37	9.65	.64	.423
PC	95.74	9.60	92.98	9.27	5.47	.021
AWM	98.12	13.20	95.58	12.72	2.47	.117
NR	91.27	12.65	89.84	11.67	.90	.342
CF	97.72	14.91	94.14	13.72	4.07	.045
PL	100.80	6.81	100.49	7.22	.12	.734
NEPSY					5.78 *	<.001
AA	9.66	2.27	8.87	2.30	7.86	.005
RS	8.81	2.61	7.48	2.50	18.21	<.001
VA	9.75	3.08	8.73	2.67	8.67	.003
TW	9.33	3.00	8.96	2.54	1.30	.256
D-KEFS					4.77 *	<.001
CWI-3	7.53	2.86	7.19	2.76	.85	.357
TMT-4	7.07	2.54	6.97	2.47	.09	.759
VFT-3	8.40	1.35	7.97	2.04	3.18	.076
CWI-4	8.06	2.09	7.93	2.21	.22	.640
DFT-3	8.36	2.35	7.97	1.87	2.10	.148
CS	9.12	2.46	7.95	2.92	10.25	.002
20Q	8.53	1.68	8.74	2.32	.58	.446
TW	9.39	2.90	9.67	2.12	.82	.367
WC	7.56	2.18	7.25	1.89	1.48	.225
DFT-2	8.87	2.22	9.68	1.84	9.82	.002

Note. AA = Auditory Attention; PC = Pair Cancellation; AWM = Auditory Working Memory; NR = Numbers Reversed; CF = Concept Formation; PL = Planning; RS = Response Set; VA = Visual Attention; TW = Tower; CWI-3 = Color Word Interference, Condition 3; TMT-4 = Trail Making Test, Condition 4; VFT-3 = Verbal Fluency Test, Condition 3; CWI-4 = Color Word Interference, Condition 4; DFT-3 = Design Fluency Test, Condition 3; CS = Card Sorting; 20Q = Twenty Questions; WC = Word Context; DFT-2: Design Fluency Test, Condition 2.

* denotes the overall multivariate F for the subscales listed below

For the D-KEFS, there was a significant multivariate effect between the two sexes, $F(10, 265) = 4.77$, $p < .001$. Further analysis showed significant differences on the Card Sorting subscale, $F(1, 274) = 10.25$, $p = .002$, and the Design Fluency Test – Condition 2 subscale, $F(1, 274) = 9.82$, $p = .002$. Females scored significantly higher ($M = 9.12$, $SD = 2.46$) than males ($M = 7.95$, $SD = 2.92$) on the Card Sorting subscale. In contrast, males scored significantly higher on the Design Fluency, Condition 2 subscale ($M = 9.68$, $SD = 1.84$) than females ($M = 8.87$, $SD = 2.22$). A marginal difference was also found on the Verbal Fluency Test, Condition 3 subscale, $F(1, 274) = 3.18$, $p = .076$), with females scoring higher ($M = 8.40$, $SD = 1.35$) than males ($M = 7.97$, $SD = 2.04$). No other significant differences were found on the other subscales, *ns.*

In order to detect the multivariate differences between language among the subscales, a series of MANOVAs were conducted (see Table 17). For the WJ III COG, there was not a significant multivariate effect between the three language categories, $F(12, 574) = 1.41$, $p = .158$. Similarly for the NEPSY and the D-KEFS, there was not a significant multivariate effect between the three language categories, $F(8, 608) = 1.54$, $p = .140$ and $F(20, 534) = .850$, $p = .656$, respectively. Although several of the individual subscales showed significant differences between language categories at the univariate level, the non-significance of the overall multivariate effects suggests that these differences may be due to Type I error. Further testing on a larger sample of participants is needed to determine whether subscale scores significantly differ by language. Table 18 shows no significant differences between ethnic categories on any of the subscales of the WJ III COG, NEPSY, or D-KEFS, all *p*'s *ns*.

Table 17

Means and Standard Deviations for Woodcock Johnson III: Tests of Cognitive Abilities (WJ III COG), NEPSY, and Delis-Kaplan Executive Function System (D-KEFS) Subtest Scores by Child's Primary Language

	English		Non-English		Not Reported			
	Mean	*SD*	Mean	*SD*	Mean	*SD*	*F*	*P*
WJ III COG							1.41 *	.158
AA	96.39	10.15	102.27	10.28	97.38	10.02	3.00	.051
PC	93.74	9.75	95.70	8.56	93.87	8.88	0.37	.690
AWM	96.69	13.14	96.95	13.43	95.40	12.12	0.29	.748
NR	90.80	11.82	91.25	16.04	88.72	11.12	0.89	.414
CF	96.21	14.75	94.09	13.84	93.18	12.39	1.32	.270
PL	100.39	7.41	104.52	7.34	100.25	5.69	3.15	.044
NEPSY							1.54 *	.140
AA	9.31	2.21	8.06	3.14	8.90	2.25	3.23	.041
RS	8.06	2.48	7.90	2.84	7.56	2.85	1.09	.337
VA	9.05	2.79	9.58	2.46	9.04	3.24	0.32	.729
TW	8.97	2.65	8.92	2.81	9.45	2.79	0.96	.385

Table 17, cont.

Means and Standard Deviations for Woodcock Johnson III: Tests of Cognitive Abilities (WJ III COG), NEPSY, and Delis-Kaplan Executive Function System (D-KEFS) Subtest Scores by Child's Primary Language

	English		Non-English		Not Reported			
	Mean	*SD*	Mean	*SD*	Mean	*SD*	*F*	*P*
D-KEFS							0.85 *	.656
CWI-3	7.27	2.74	7.18	3.31	7.45	2.85	0.13	.882
TMT-4	7.11	2.49	6.44	2.66	6.90	2.41	0.69	.501
VFT-3	8.28	1.87	7.50	2.58	7.79	1.56	2.92	.056
CWI-4	8.05	2.15	7.39	2.17	7.95	2.22	0.77	.465
DFT-3	8.07	2.06	8.53	2.15	8.02	1.91	0.47	.627
CS	8.41	2.99	7.38	1.76	8.31	2.60	1.09	.338
20Q	8.78	2.11	7.90	2.10	8.62	2.20	1.43	.241
TW	9.60	2.29	10.00	2.17	9.38	2.73	0.52	.593
WC	7.34	2.04	7.11	2.00	7.41	1.84	0.16	.855
DFT-2	9.43	1.97	10.01	2.96	9.27	1.74	0.99	.372

Note. AA = Auditory Attention; PC = Pair Cancellation; AWM = Auditory Working Memory; NR = Numbers Reversed; CF = Concept Formation; PL = Planning; RS = Response Set; VA = Visual Attention; TW = Tower; CWI-3 = Color Word Interference, Condition 3; TMT-4 = Trail Making Test, Condition 4; VFT-3 = Verbal Fluency Test, Condition 3; CWI-4 = Color Word Interference, Condition 4; DFT-3 = Design Fluency Test, Condition 3; CS = Card Sorting; 20Q = Twenty Questions; WC = Word Context; DFT-2; Design Fluency Test, Condition 2.
* denotes the overall multivariate F for the subscales listed below

Table 18

Means and Standard Deviations for Woodcock Johnson III: Tests of Cognitive Abilities (WJ III COG), NEPSY, and Delis-Kaplan Executive Function System (D-KEFS) Subtest Scores by Child's Ethnicity

	Caucasian		African American		Hispanic/ Latino American		Not Reported			
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	*F*	*P*
WJ III COG									1.20 *	.255
AA	96.23	10.81	94.37	7.02	101.05	8.27	97.07	10.35	1.66	.176
PC	94.04	9.54	91.41	5.30	93.80	7.47	94.10	9.65	0.24	.867
AWM	99.38	12.56	92.03	10.93	97.84	13.01	95.69	13.13	1.69	.169
NR	91.90	11.65	86.52	9.43	87.31	11.86	90.81	11.97	1.29	.279
CF	96.82	13.11	85.52	9.57	94.07	12.54	96.01	14.68	1.85	.139
PL	100.07	7.59	99.26	8.02	102.45	6.47	100.46	7.00	0.82	.485
NEPSY									1.07 *	.384
AA	9.42	2.06	8.12	2.31	9.33	2.70	9.04	2.35	1.10	.349
RS	8.34	2.58	6.69	3.66	7.76	3.22	7.84	2.48	1.37	.251
VA	9.59	3.40	7.95	3.11	9.05	2.59	9.03	2.74	1.17	.323
TW	8.87	2.80	7.49	3.49	8.84	2.81	9.30	2.53	1.86	.137

Table 18, cont.

Means and Standard Deviations for Woodcock Johnson III: Tests of Cognitive Abilities (WJ III COG), NEPSY, and Delis-Kaplan Executive Function System (D-KEFS) Subtest Scores by Child's Ethnicity

	Caucasian		African American		Hispanic/ Latino American		Not Reported			
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	*F*	*P*
D-KEFS									0.79 *	.782
CWI-3	7.90	2.64	7.50	3.91	6.98	3.04	7.10	2.73	1.26	.288
TMT-4	7.10	2.79	6.47	3.35	6.87	2.76	7.10	2.32	0.24	.871
VFT-3	8.38	1.57	7.27	1.47	7.71	1.68	8.09	1.96	1.35	.259
CWI-4	8.32	2.24	8.06	2.45	7.49	2.30	7.94	2.14	0.85	.470
DFT-3	8.17	2.20	7.75	1.78	7.62	1.93	8.05	1.98	0.47	.703
CS	8.96	3.22	9.48	4.22	7.50	2.64	8.23	2.66	2.10	.101
20Q	9.28	2.24	8.75	2.70	8.29	1.93	8.60	2.09	1.75	.156
TW	9.97	2.41	9.43	3.99	9.51	2.06	9.49	2.37	0.58	.629
WC	7.81	2.26	7.43	2.00	6.84	2.12	7.29	1.89	1.51	.213
DFT-2	9.66	2.07	8.79	2.04	9.01	2.63	9.47	1.88	0.91	.436

Note. AA = Auditory Attention; PC = Pair Cancellation; AWM = Auditory Working Memory; NR = Numbers Reversed; CF = Concept Formation; PL = Planning; RS = Response Set; VA = Visual Attention; TW = Tower; CWI-3 = Color Word Interference, Condition 3; TMT-4 = Trail Making Test, Condition 4; VFT-3 = Verbal Fluency Test, Condition 3; CWI-4 = Color Word Interference, Condition 4; DFT-3 = Design Fluency Test, Condition 3; CS = Card Sorting; 20Q = Twenty Questions; WC = Word Context; DFT-2; Design Fluency Test, Condition 2.
* denotes the overall multivariate F for the subscales listed below

In order to detect the multivariate differences between diagnosis and the varying subscales of the WJ III COG, NEPSY and D-KEFS, a series of MANOVAs were conducted (see Tables 19 and 20). For the WJ III COG, there was a significant multivariate effect between the diagnoses, $F(30, 1370) = 5.06$, $p < .001$. Further analysis showed significant differences on the Pair Cancellation, $F(5, 275) = 3.44$, $p = .005$, Auditory Working Memory, $F(5, 275) = 5.63$, $p < .001$, Numbers Reversed $F(5, 275) = 8.21$, $p < .001$, Concept Formation $F(5, 275) = 3.98$, $p = .002$, and Planning subscales $F(5, 275) = 17.94$, $p < .001$.

Least Square Difference (LSD) post hoc comparisons showed that on the Pair Cancellation subtest, children with ADHD/ ADD ($M = 90.73$, $SD = 7.73$) had significantly lower scores than individuals with a learning disability ($M = 95.48$, $SD = 8.26$) or language disability ($M = 99.79$, $SD = 6.10$). Furthermore, individuals with ADHD/ADD scored significantly lower than individuals without reported diagnostic classifications ($M = 95.43$, $SD = 10.25$).

On Auditory Working Memory, individuals with ADHD/ ADD ($M = 88.80$, $SD = 9.84$) scored significantly lower than individuals with a learning disability ($M = 97.37$, $SD = 12.31$), language disability ($M = 100.09$, $SD = 11.09$), ASD ($M = 99.44$, $SD = 11.88$), or unreported diagnosis ($M = 100.31$, $SD = 12.93$). Furthermore, children with multiple disabilities ($M = 92.79$, $SD = 13.11$) scored significantly lower than children with a learning disability, ASD, or unknown diagnosis.

Table 19

Means and Standard Deviations for Woodcock Johnson III: Tests of Cognitive Abilities (WJ III COG), NEPSY, and Delis-Kaplan Executive Function System (D-KEFS) Subtest Scores by Child's Diagnosis

	Learning Disability		Language Disability		ADHD/ ADD		Autism Spectrum Disorder		Multiple Disabilities		Not Reported	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
WJ III COG												
AA	98.20	9.55	104.88	9.34	95.71	95.11	95.11	11.25	97.61	11.44	96.17	8.17
PC	95.48	8.26	99.79	6.10	90.73	7.73	94.24	8.57	91.50	10.25	95.43	10.25
AWM	97.37	12.31	100.09	11.09	88.80	9.84	99.44	11.88	92.79	13.11	100.31	12.93
NR	89.60	9.86	100.50	4.92	82.69	8.68	94.06	11.82	88.22	13.20	95.04	11.65
CF	99.55	11.98	99.95	8.95	89.85	10.26	97.48	13.20	91.64	15.59	97.12	16.38
PL	101.88	4.10	115.00	0.00	102.18	4.15	100.33	7.06	97.27	8.76	100.55	3.89
NEPSY												
AA	9.41	2.62	8.82	3.71	8.94	2.27	9.74	2.06	8.67	2.07	9.20	1.79
RS	7.83	2.45	9.64	1.63	6.70	2.90	8.74	2.96	8.16	2.34	7.44	2.57
VA	9.03	3.07	10.00	2.76	7.89	2.79	9.77	3.10	9.21	2.79	9.22	2.48
TW	8.99	2.56	10.31	1.49	8.37	2.66	9.43	2.48	8.86	3.00	9.72	2.46

Table 19, cont.

Means and Standard Deviations for Woodcock Johnson III: Tests of Cognitive Abilities (WJ III COG), NEPSY, and Delis-Kaplan Executive Function System (D-KEFS) Subtest Scores by Child's Diagnosis

	Learning Disability		Language Disability		ADHD/ ADD		Autism Spectrum Disorder		Multiple Disabilities		Not Reported	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
D-KEFS												
CWI-3	6.79	1.95	9.18	1.17	6.18	2.80	7.65	1.77	7.96	2.99	6.00	3.29
TMT-4	7.17	1.78	7.91	1.51	6.66	2.66	7.69	2.48	6.48	2.75	7.10	2.20
VFT-3	9.38	1.76	6.27	0.13	7.08	1.43	8.10	1.85	7.74	1.71	7.77	1.43
CWI-4	7.52	2.12	8.64	0.81	7.32	1.66	8.62	1.32	8.00	2.27	7.80	2.73
DFT-3	8.22	1.38	10.00	0.00	8.09	1.57	8.62	1.57	7.58	2.17	8.68	1.68
CS	7.17	2.53	8.00	0.00	8.68	0.71	10.28	1.67	7.81	3.78	8.83	2.59
20Q	8.87	1.09	6.00	0.00	8.00	1.07	9.18	2.88	9.67	2.71	8.02	0.77
TW	10.32	1.58	11.45	0.27	9.09	1.32	9.00	2.37	9.54	2.97	10.03	1.85
WC	8.07	2.60	7.23	0.81	5.72	0.62	6.98	1.85	7.86	1.62	7.30	1.76
DFT-2	9.62	1.64	13.00	0.00	8.54	1.66	11.38	0.79	8.70	1.40	9.06	1.56

Note. ADHD = Attention Deficit Hyperactivity Disorder; ADD = Attention Deficit Disorder; AA = Auditory Attention; PC = Pair Cancelation; AWM = Auditory Working Memory; NR = Numbers Reversed; CF = Concept Formation; PL = Planning; RS = Response Set; VA = Visual Attention; TW = Tower; CWI-3 = Color Word Interference, Condition 3; TMT-4 = Trail Making Test, Condition 4; VFT-3 = Verbal Fluency Test, Condition 3; CWI-4 = Color Word Interference, Condition 4; DFT-3 = Design Fluency Test, Condition 3; CS = Card Sorting; 20Q = Twenty Questions; WC = Word Context; DFT-2: Design Fluency Test, Condition 2.

Table 20

F and P Values for Woodcock Johnson III: Tests of Cognitive Abilities (WJ III COG), NEPSY, and Delis-Kaplan Executive Function System (D-KEFS) Subtest Scores by Diagnosis

TEST/ Subtest	*F*		*P*
WJ III COG	5.06	*	<.001
Auditory Attention	2.07		.069
Pair Cancellation	3.44		.005
Auditory Working Memory	5.63		<.001
Numbers Reversed	8.21		<.001
Concept Formation	3.98		.002
Planning	17.94		<.001
NEPSY	2.31	*	<.001
Auditory Attention	1.57		.167
Response Set	3.92		.002
Visual Attention	1.91		.093
Tower	1.84		.106
D-KEFS	9.44	*	<.001
Color Word Interference 3	5.90		<.001
Trail Making Test 4	1.89		.097
Verbal Fluency 3	13.98		<.001
Color Word Interference 4	2.04		.073
Design Fluency 3	5.49		<.001
Card Sorting	7.03		<.001
Twenty Questions	10.14		<.001
Tower	3.93		.002
Word Context	7.97		<.001
Design Fluency 2	33.91		<.001

* denotes the overall multivariate F for the subscales listed below

On the Numbers Reversed subtest, children with ADHD/ ADD (M = 82.69, SD = 8.68), performed significantly worse than children with a learning disability (M = 89.60, SD = 9.86), language disability (M = 100.50, SD = 4.92), ASD (M = 94.06, SD = 11.82), or the unknown diagnosis group (M = 95.04, SD = 11.65). Children with multiple disabilities scored significantly worse than children with a language disability, ASD, or unknown diagnosis. Furthermore, on the Numbers Reversed subscale, children with a learning disability scored lower than children with a language disability and children with an ASD. On the Concept Formation subtest children with ADHD/ ADD ($M = 89.85$, $SD = 10.26$) performed significantly worse than children with a learning disability ($M = 99.55$, $SD = 11.98$), language disability ($M = 99.95$, $SD = 8.95$), ASD ($M = 97.48$, $SD = 13.20$), or unknown diagnosis ($M = 97.12$, $SD = 16.38$). Children with multiple disabilities ($M = 91.64$, $SD = 15.59$) performed significantly lower than children with a learning disability, ASD, or unknown diagnosis. On the Planning subtest, children with a language disability ($M = 115.00$, $SD = 0.00$) performed significantly better than every other group (*M*s = 97.27-102.18, *SD*s = 3.89-8.76). This may be due to the restricted range for the language disability scores on this subtest. Also on the Planning subtest, children with multiple disabilities ($M = 97.27$, $SD = 8.76$) performed significantly worse than ever other group. No significant differences were found on the Auditory Attention subtest, *ns*.

For the NEPSY, there was a significant multivariate effect between the diagnosis categories, $F(20, 1144) = 2.31$, $p < .001$. Further analysis showed significant differences on the Response Set subscale, $F(5, 286) = 3.92$, $p = .002$. On this subscale, children with ADHD/ ADD ($M = 6.70$, $SD = 2.90$) scored significantly lower than children with a learning disability ($M = 7.83$, $SD = 2.45$), language disability ($M = 9.64$, $SD = 1.63$), ASD ($M = 8.74$, $SD = 2.96$), or multiple disability ($M = 8.16$, $SD = 2.34$). Children with an unknown diagnosis ($M = 7.44$, $SD = 2.57$) performed significantly worse than children with a language disability or an ASD. Furthermore, children with language disability performed significantly better than children with a learning disability. No other significant differences were found on the other subscales, *ns*.

For the D-KEFS, there was a significant multivariate effect between the diagnosis groups, $F(50, 1245) = 9.44$, $p < .001$. Further analysis showed significant differences on the Color Word Interference, Condition 3, $F(5, 254) = 5.90$, $p < .001$, the Verbal Fluency Test, Condition 3, $F(5, 254) = 13.98$, $p < .001$, the Design Fluency Test, Condition 3 $F(5, 254) = 5.49$, $p < .001$, the Card Sorting $F(5, 254) = 7.03$, $p < .001$, the Twenty Questions $F(5, 254) = 10.14$, $p < .001$, the Tower $F(5, 254) = 3.93$, $p = .002$, the Word Context $F(5, 254) = 7.97$, $p < .001$, and the Design Fluency Test, Condition 2 tests $F(5, 254) = 33.91$, $p < .001$.On Color Word Interference Test, Condition 3, children with unreported diagnoses ($M = 6.00$, $SD = 3.29$), scored significantly worse than children with a language disability ($M = 9.18$, $SD = 1.17$), ASD ($M = 7.65$, $SD = 1.77$), or multiple disabilities ($M = 7.96$, $SD = 2.99$). Children with ADHD/ ADD ($M = 6.18$, $SD = 2.80$) performed significantly worse than children with a language disability, ASD, or multiple disabilities. Subjects with a learning disability ($M = 6.79$, $SD = 1.95$) diagnosis performed significantly worse than subjects with a language disability or multiple disabilities.

On the Verbal Fluency Test, Condition 3, children with a learning disorder diagnosis ($M = 9.38$, $SD = 1.76$) performed significantly better than every other group (Ms= 6.27-8.10, SDs = 0.13-1.85). Children with a language disorder ($M = 6.27$, SD =

0.13) diagnosis performed significantly worse than every group (*M* s= 7.74-9.38, *SD* = 1.43-1.85), with the exception of students with ADHD/ ADD. Children with an ASD (*M* = 8.10, *SD* = 1.85) performed significantly better than children with ADHD/ ADD (*M* = 7.08, *SD* = 1.43).

On the Design Fluency Test, Condition 3, children with a language disability diagnosis (*M* = 10.00, *SD* = 0.00) performed significantly better than any other group. However, this may be due to the restriction in range and Type I error for students with a language disability on this subtest. Further testing on a larger sample of participants is needed to determine whether the performance of children with a language disability is significantly discrepant from children with other clinical disorders on this measure. Children with multiple disabilities (*M* = 7.58, *SD* = 2.17) scored significantly worse than students with a language disability, learning disability (*M* = 8.22, *SD* = 1.38), ASD (*M* = 8.62, *SD* = 1.57), or unreported diagnosis (*M* = 8.68, *SD* = 1.68). On the Card Sorting test, children with an ASD (*M* = 10.28, *SD* = 1.67) performed significantly better than every other group (*Ms* = 7.17-8.83, *SD* = 0.00-3.78). Students with a learning disability (*M* = 7.17, *SD* = 2.53) also scored worse than students with ADHD/ ADD (*M* = 8.68, *SD* = 0.71) and unreported disorders (*M* = 8.83, *SD* = 2.59).

On the Twenty Questions test, students with a language disorder (*M* = 6.00, *SD* = 0.00) performed worse than every other group; however, this may be due to the restriction in range. Children with ADHD/ ADD (*M* = 8.00, *SD* = 1.07) performed significantly worse than students with a learning disability (*M* = 8.87, *SD* = 1.09), ASD

(*M* = 9.18, *SD* = 2.88), and multiple disabilities (*M* = 9.67, *SD* = 2.71). Students with multiple disabilities scored significantly better than students with a learning disability or unreported disability (*M* = 8.02, *SD* = 0.77). Furthermore, students with multiple disabilities scored significantly better than students with an ASD.

On the D-KEFS Tower test, children with a language disability (*M* = 11.45, *SD* = 0.27) or a learning disability (*M* = 10.32, *SD* = 1.58) performed significantly better than children with ADHD/ ADD (*M* = 9.09, *SD* = 1.32), ASD (*M* = 9.00, *SD* = 2.37), or multiple disabilities (*M* = 9.54, *SD* = 2.97). Children with an unreported diagnosis (*M* = 10.03, *SD* = 1.85) scored significantly worse than students with ASD. On the Word Context test, students with ADHD/ ADD (*M* = 5.72, *SD* = 0.62) performed significantly worse than every other group (*M* s= 6.98-8.07, *SD*s= 0.81-2.60). Children with an ASD (*M* = 6.98, *SD* = 1.85) performed significantly worse than children with a learning disability (*M* = 8.07, *SD* = 2.60) or multiple disabilities (*M* = 7.86, *SD* = 1.62). Students with a learning disability performed significantly better than students with an unreported diagnosis (*M* = 7.30, *SD* = 1.76).

On the Design Fluency Test, Condition 2, children with a language disorder diagnosis (*M* = 13.00 *SD* = 0.00) performed significantly better than every other group; however, this may be due to restriction in range. Students with an ASD (*M* = 11.38, *SD* = 0.79) performed significantly better than every group, with the exception of students with a language disability. Subjects with a learning disability (*M* = 9.62, *SD* = 1.64) performed significantly better than students with ADHD/ ADD (*M* = 8.54, *SD* = 1.66), multiple disabilities ($M = 8.70$, $SD = 1.40$), or with an unreported diagnosis ($M = 9.06$, $SD = 1.56$). No other significant differences were found on the other subscales, *ns*.

The results of the MANOVAs conducted on this sample revealed a significant effect on subtest scores when either gender or disability criteria were taken into account. Regarding the disability criteria category, the samples of students with ADHD or LD appeared to have a large effect on subtest scores. Although other diagnostic categories, such as the language disorder group, had significant effect on scores, these groups had smaller overall *N*s than the ADHD and LD groups. Therefore, in order to account for any confounds, for the primary analysis, confirmatory factor analyses will be conducted using four separate groups: the full sample, a sample that includes only boys, a sample with students diagnosed with a learning disability removed, and a sample with students with ADHD removed.

Primary Analysis

Confirmatory Factor Analysis

For each significant model, results are displayed as a path diagram, which is a graphical representation of a hypothesized causal model. A path diagram consists of paths along which causal relations are presented. The paths are designated with single-direction arrows that represent a direct relationship between a variable and factor. Path diagrams contain observed variables (e.g., subtests/tests), which are indicated by rectangles, and unobserved/latent variables (e.g., factors), which are indicated by ovals.

Additionally, values, also referred to as path coefficients, are assigned to each arrow quantifying how well that variable loads onto the corresponding factor. The loadings are represented as standardized estimates. Typically, path coefficients range from 0.0-1.0; however, in some cases a path coefficient may be greater than 1.0, which typically indicates a potential for high level of multicollinearity among the indicators. Such results are discussed in more detail throughout the remainder of this chapter. Path coefficients were reported to show the relative magnitude of each path in a comparable number, with higher values indicating a stronger causal relationship. Prior to running the confirmatory factor analyses, a constraint value of 1 was placed on one measure/observed variable of each latent factor. This type of constraint is common in model estimations containing variables that have a defined scale (Shumacker & Lomax, 2004), which is the case with this study where all of the measured variables are standardized tests.

In addition to the path diagrams, each model resulted in goodness-of-fit statistics, which are measurements that indicate how well the factor structure explained the data set that was being examined (Keith, 2005). The fit statistics used in the present study included the adjusted chi-square test (χ^2, cut off value > 3), the root mean square error of approximation (RMSEA, cut off value < .09), the comparative fit index (CFI, cut off > .70), the Non-Normed Fit Index (NNFI, cut off value >.70), and a standardized root square mean residual (SRMR, cut off value < .09). Furthermore, a single-sample cross-validation index (ECVI) was calculated for each of the models. The alternative model with the smallest ECVI value is often the most stable in the population. (Browne & Cudeck, 1993).

Research question. Is the underlying factor structure of executive functioning tasks across the WJ III COG, NEPSY, and D-KEFS in a clinical population of children best described by:

a. A theoretical model in which all tasks load on one, general executive functioning factor?

b. Anderson and colleagues' (2002) theory of executive functioning?

c. The CHC model of cognitive abilities (McGrew, 2005)?

d. The School Neuropsychological conceptual model (SNP model; Miller, 2007)?

Four separate confirmatory factor analyses were conducted with corresponding models of increasing complexity. The models were then compared to determine which model best fit the data set and best described the conceptual process of executive functioning. Confirmatory factor analysis was conducted with the LISREL 8.8 statistical program.

Model one (one-factor model). **Is the underlying factor structure of executive functioning tasks across the WJ III COG, NEPSY, and D-KEFS in a clinical population of children best described by a theoretical model in which all tasks load on one, general executive functioning factor?** This model was proposed as the "base" model to be compared with the more complex multifactor models. This simple, one-factor model serves as a comparison tool in contrast to more advanced models. It is desirable to rule out one-factor models so that support for multifactor models is stronger (Thompson, 2004).

The path diagrams with factor loadings of all the subtests are presented in Table 21. As mentioned, factor loadings are values that describe how well a specific variable measures a factor (Tabachnick & Fidell, 2001). Higher loading values indicate that a variable is a more pure measure of a factor. According to Comrey and Lee (1992), loadings of .71 and greater are considered "excellent," loadings between .63 and .70 are considered "very good," .55 to .62 are considered "good," .45 to .54 are considered "fair," and .32 to .44 are considered "poor." Tabachnick and Fidell suggest that only loadings of .32 and higher are interpretable.

The majority of the subscales across all three batteries loaded poorly on the one-factor model of executive functioning, with the WJ III COG Planning, D-KEFS Verbal Fluency Test, Condition 3, D-KEFS Twenty Questions, and D-KEFS Tower subtests loading the weakest (.14, .09, .10, and .11 respectively) in the full sample. The WJ III COG Pair Cancelation, Auditory Working Memory, Numbers Reversed, and Concept Formation subtests had the strongest factor loadings ranging from .58 to .68 in the full sample. Similar trends in relative strength of loadings was found across the full sample and subsamples of boys only, individuals with learning disabilities removed, and individuals with ADHD removed. Though each subsample demonstrated similar patterns of strengths, the subsample with learning disorders removed and the subsample with ADHD removed had slightly stronger loadings on the four strongest loadings.

Table 21

Path Coefficients of Model One Including the Full Sample and Three Sub-Samples

	Full Model	Boys Only	No LD	No ADHD
Executive Functioning				
NPRS	.40 $*_5$	.38 $*_5$	.33 $*_5$	.46 $*_4$
NPVA	.31 $*_6$	.35 $*_5$	.27 $*_6$	.36 $*_5$
NPTW	.36 $*_5$	.34 $*_5$	.36 $*_5$	.32 $*_5$
WJAA	.31 $*_6$	.24 $*_6$	.34 $*_5$	.30 $*_6$
WJPC	.58 $*_3$	.56 $*_3$	.60 $*_3$	.59 $*_3$
WJAWM	.62 $*_3$	.55 $*_3$	.65 $*_2$	.59 $*_3$
WJNR	.68 $*_2$	.61 $*_3$	.69 $*_2$	.65 $*_2$
WJCF	.67 $*_2$	.62 $*_3$	.70 $*_2$	.65 $*_2$
WJPL	.14 $*_6$	.15	.12	.12
DKCW3	.28 $*_6$	.27 $*_6$	.25 $*_6$	.29 $*_6$
DKTMT4	.33 $*_5$	.41 $*_5$	.37 $*_5$	.31 $*_6$
DKVF3	.09	.10	.06	.09
DKCW4	.36 $*_5$	.43 $*_5$	.32 $*_5$	.37 $*_5$
DKDF3	.21 $*_6$	.21 $*_6$	.21 $*_6$	.22 $*_6$
DKCS	.27 $*_6$	.29 $*_6$	.32 $*_5$	.29 $*_6$
DK20Q	.10	.16 $*_6$	.08	.18 $*_6$
DKTOW	.11	.05	.09	.11
DKWC	.26 $*_6$	.28 $*_6$	.23 $*_6$	.30 $*_6$
DKDF2	.31 $*_6$	.38 $*_5$	.34 $*_5$	.31 $*_6$

Note. LD = Learning Disability; ADHD = Attention Deficit Hyperactivity Disorder; NPRS = NEPSY Response Set; NPVA = NEPSY Visual Attention; NPTW = NEPSY Tower; WJAA = WJ III COG Auditory Attention; WJPC = WJ III COG Pair Cancellation; WJAWM = WJ IIICOG Auditory Working Memory; WJNR = WJ III COG Numbers Reversed; WJCF = WJ III COG Concept Formation; DKCW3 = D-KEFS Color Word Interference, Condition 3; DKTMT4 = D-KEFS Trail Making Test, Condition 4; DKVF3 = D-KEFS Verbal Fluency, Condition 3; DKCW4 = D-KEFS Color Word Interference, Condition 4; DKDF3 = D-KEFS Design Fluency, Condition 3; DKCS = D-KEFS Card Sorting; DK20Q = D-KEFS Twenty Questions; DKTOW = D-KEFS Tower; DKWC = D-KEFS Word Context; DKDF2 = D-KEFS Design Fluency, Condition 2.

* $p < .05$. Factor loadings are noted by the following subscripts: 1 = "Excellent"; 2 = "Very Good"; 3 = "Good"; 4 = "Fair"; 5 = "Poor"; 6 = "Uninterpretable."

The fit statistics for the one-factor model of executive functioning are displayed in Table 22. Adjusted chi square tests across all subsamples resulted in values greater than 3, suggesting that a one-factor solution adequately fits the model. In terms of ascertaining model fit, an adjusted chi square is rarely used in isolation. For the full sample and the four subsamples, the RMSEA and SRMR fit indices were not less than .09, indicating a lack of fit of the model. In addition, across all subsamples and the full sample, CFIs and NNFIs were all below the critical value of .7, indicating that the overall one-factor model of executive functioning is not a good fit.

Model two (Anderson et al.'s model). **Is the underlying factor structure of executive functioning tasks across the WJ III COG, NEPSY, and D-KEFS in a clinical population of children best described by Anderson and colleagues' (2002) theory of executive functioning?** Anderson and colleagues operationalized executive functions to factor into the components of attentional control, goal setting, and cognitive flexibility. This model focuses specifically on executive functioning, while also encompassing a wide range of executive functioning subcomponents.

The path diagrams with factor loadings on all subtests for this model are presented in Table 23. For the first factor, attentional control, factor loadings ranged from .25 (WJ III COG Auditory Attention) to .56 (WJ III COG Pair Cancellation) in the full sample. The subsample excluding individuals with learning disorders, had lower path coefficients for the following subscales: NEPSY Auditory Attention and NEPSY Response Set. The subsample excluding individuals with learning disorders and the subsample excluding individuals with ADHD had higher path coefficients on the WJ III COG Auditory Attention subscale.

Table 22

Fit Indicies of the Four Proposed Models Including the Full Sample and Three Sub-Samples

	Adj Chi^2	RMSEA	SRMR	NNFI	CFI	ECVI
Model 1						
Full Sample	4.323	.104	.094	.600	.644	2.380
Boys Only	4.433	.127	.112	.424	.488	3.519
LD Removed	4.167	.112	.103	.531	.583	2.866
ADHD Removed	4.165	.108	.098	.575	.622	2.612
Model 2						
Full Sample	4.499	.102	.099	.584	.634	2.560
Boys Only	4.488	.121	.118	.412	.484	3.629
LD Removed	4.207	.110	.119	.512	.571	3.091
ADHD Removed	4.117	.102	.100	.585	.635	2.680
Model 3 with WJAA						
Full Sample	4.069	.096	.092	.635	.714	2.230
Model 3 without WJAA						
Full Sample	3.995	.095	.090	.653	.722	2.020
Boys Only	3.978	.113	.104	.520	.616	2.899
LD Removed	3.677	.103	.109	.600	.679	2.489
ADHD Removed	3.681	.097	.093	.655	.724	2.200
Model 4						
Full Sample	3.185	.078	.078	.739	.796	1.452
Boys Only	3.457	.099	.098	.568	.661	2.216
LD Removed	2.951	.086	.086	.690	.757	1.819
ADHD Removed	3.555	.090	.087	.662	.735	1.804

Note. Adj Chi^2 = adjusted chi-square; RMSEA = root mean square error of approximation; SRMR = standardized root mean squared residual; NNFI = Non-Normed Fit Index; CFI = comparative fit index; ECVI = single-sample cross validation index; LD = Learning Disability; ADHD = Attention Deficit Hyperactivity Disorder; WJAA = WJ III COG Auditory Attention.

Table 23

Path Coefficients of Model Two Including the Full Sample and Three Sub-Samples

	Full Model	Males Only	No LD	No ADHD
Attentional Control				
NPAA	.44 $*_5$	.42 $*_5$	.25 $*_6$	.41 $*_5$
NPRS	.54 $*_4$	.51 $*_4$	.37 $*_5$	.53 $*_4$
NPVA	.37 $*_5$	.38 $*_5$	.30 $*_6$	.36 $*_5$
WJAA	.25 $*_6$	.22 $*_6$	.39 $*_5$	.30 $*_6$
WJPC	.56 $*_3$	.59 $*_3$	.63 $*_2$	.58 $*_3$
DKCW3	.32 $*_5$	.27 $*_6$	.26 $*_6$	.28 $*_6$
DKDF2	.28 $*_6$	.30 $*_6$	.29 $*_6$	.26 $*_6$
Cognitive Flexibility				
WJAWM	.64 $*_2$	.61 $*_3$	.68 $*_2$	.61 $*_3$
WJNR	.70 $*_2$	.68 $*_2$	.71 $*_1$	.68 $*_2$
WJCF	.69 $*_2$	.66 $*_2$	.72 $*_1$	.67 $*_2$
DKTMT4	.30 $*_6$	.34 $*_5$	.32 $*_5$	.28 $*_6$
DKVF3	.11 $*_6$	.10 $*_6$	.07	.09
DKCW4	.34 $*_5$	.35 $*_5$	.30 $*_6$	.36 $*_5$
DKDF3	.16 $*_6$	.13 $*_6$	.17 $*_6$	.18 $*_6$
Goal Setting				
NPTW	.22 $*_6$	.17 $*_6$	.08 $*_6$	.14 $*_6$
WJPL	.00	.00	.14 $*_6$	.00
DKCS	.61 $*_3$	.81 $*_1$	-.31 $*_6$	.55 $*_3$
DK20Q	.37 $*_5$	.34 $*_5$	.33 $*_5$	.42 $*_5$
DKTOW	.22 $*_6$	.28 $*_6$	.13 $*_6$	.28 $*_6$
DKWC	.40 $*_5$	.33 $*_5$	-.03	.61 $*_3$

Note. LD = Learning Disability; ADHD = Attention Deficit Hyperactivity Disorder; NPRS = NEPSY Response Set; NPVA = NEPSY Visual Attention; NPTW = NEPSY Tower; WJAA = WJ III COG Auditory Attention; WJPC = WJ III COG Pair Cancellation; WJAWM = WJ IIICOG Auditory Working Memory; WJNR = WJ III COG Numbers Reversed; WJCF = WJ III COG Concept Formation; DKCW3 = D-KEFS Color Word Interference, Condition 3; DKTMT4 = D-KEFS Trail Making Test, Condition 4; DKVF3 = D-KEFS Verbal Fluency, Condition 3; DKCW4 = D-KEFS Color Word Interference, Condition 4; DKDF3 = D-KEFS Design Fluency, Condition 3; DKCS = D-KEFS Card Sorting; DK20Q = D-KEFS Twenty Questions; DKTOW = D-KEFS Tower; DKWC = D-KEFS Word Context; DKDF2 = D-KEFS Design Fluency, Condition 2.

* $p < .05$. Factor loadings are noted by the following subscripts: 1 = "Excellent"; 2 = "Very Good"; 3 = "Good"; 4 = "Fair"; 5 = "Poor"; 6 = "Uninterpretable."

Within the second factor, cognitive flexibility, coefficients ranged from .11 (D-KEFS Verbal Fluency, Condition 3) to .70 (WJ III COG Numbers Reversed) in the full sample. The strongest factor loadings were with the WJ III COG Auditory Working Memory, Concept Formation, and Numbers Reversed subtests (.64, .69, .70, respectively). A similar pattern occurred across all the subsamples. When compared to the other two hypothesized factors, the final hypothesized factor, goal setting, had lower coefficient values overall, ranging from < .01 (WJ III COG Planning) to .61 (D-KEFS Card Sorting). Across all subscale scores, the subsample that excluded individuals with learning disorders had lower factor loadings.

The fit statistics for the two-factor model of executive functioning are displayed in Table 22. Adjusted chi square tests across all subsamples resulted in values greater than 3, suggesting that a three-factor solution adequately fits the model. However, for the full sample, the RMSEA and SRMR values were not less than .09, indicating a lack of fit of the model. In addition, all three subsamples failed to cross the RMSEA critical value, with values ranging from .100-.119. Across all subsamples and the full sample, NNFIs and CFIs were all below the critical value of .7, indicating that the overall three-factor model of executive functioning was not a good fit to the utilized data.

Model three (CHC Theory model). **Is the underlying factor structure of executive functioning tasks across the WJ III COG, NEPSY, and D-KEFS in a clinical population of children, best described by the CHC model of cognitive abilities?** The CHC theory (McGrew, 2005) asserts that a three-tier hierarchical model of cognitive abilities can explain cognitive functioning. The CHC model was analyzed due to its support in the literature as well as its influence on cross-battery assessment, which is an important component of neuropsychological evaluations. Within CHC theory, the second tier includes broad abilities that represent general cognitive abilities. The broad cognitive abilities included in the current study are Comprehension-Knowledge (*Gc*), Long-Term Retrieval (*Glr*), Visual-Spatial Thinking (*Gv*), Fluid Reasoning (*Gf*), Processing Speed (*Gs*), and Short-Term Memory (*Gsm*).

Although a seventh broad ability, Auditory Processing (*Ga*) is a core component of the CHC theory, only one of the subtests utilized in the current study (WJ III COG Auditory Attention) could be conceptualized to load on the *Ga* factor. Including a factor with only one path is less than ideal in confirmatory factor analysis (Tabachnick & Fidell, 2001). However, due to the significance of the *Ga* factor within the CHC theoretical framework, model 3 was first analyzed with the *Ga* factor. While the model including the *Ga* factor was a good fit in the full sample (See Table 22), path coefficients could not be converged for the full sample or any of the subsamples, therefore the single measure factor of *Ga* was removed from model three.

The path diagrams with factor loadings on all subtests for this model are presented in Table 24. For the first factor, *Gsm*, the WJ III COG Auditory Working Memory and WJ III COG Numbers Reversed subtests had very good to excellent factor loadings (.64 & .71, respectively). However, the NEPSY Auditory Attention and Response Set tests loaded poorly (.37 & .44, respectively) in the full sample.

Table 24

Path Coefficients of Model Three Including the Full Sample and Three Sub-Samples

	Full Model	Males Only	No LD	No ADHD
Gsm				
NPAA	.37 $*_5$	.37 $*_5$	.24 $*_6$	.44 $*_5$
NPRS	.44 $*_5$	.41 $*_5$	.34 $*_5$	.52 $*_4$
WJAWM	.64 $*_2$	.63 $*_2$	.68 $*_2$	.62 $*_3$
WJNR	.71 $*_1$	.69 $*_2$	.71 $*_1$	.69 $*_2$
Gs				
WJPC	.65 $*_2$	.59 $*_3$	.57 $*_3$	.61 $*_3$
DKTMT4	.36 $*_5$	.39 $*_5$	.41 $*_5$	.34 $*_5$
Gc				
DKWC	.39 $*_5$	.27 $*_6$	-	.93 $*_1$
DK20Q	.37 $*_5$	.45 $*_4$	-	.31 $*_6$
Glr				
DKCW3	.56 $*_3$	.45 $*_4$	.60 $*_3$	.57 $*_3$
DKVF3	.12	.08	.07	.09
DKCW4	.88 $*_1$	.93 $*_1$	.89 $*_1$	.83 $*_1$
Gf				
WJCF	.46 $*_4$	.38 $*_5$	.67 $*_2$	.57 $*_3$
DKCS	.24 $*_6$	.25 $*_6$	.30 $*_6$	.24 $*_6$
Gv				
WJGPL	.17	.25	.20 $*_6$	.20
NPVA	.21 $*_6$	.34 $*_5$	.23 $*_6$	.15 $*_6$
NPTW	.20 $*_6$	.20 $*_6$	.24 $*_6$	.17
DKDF3	.59 $*_3$	.62 $*_3$	.64 $*_2$	.58 $*_3$
DKTOW	.22 $*_6$	.06	.17	.76 $*_1$
DKDF2	.73 $*_1$	.58 $*_3$	.72 $*_1$	.28

Note. LD = Learning Disability; ADHD = Attention Deficit Hyperactivity Disorder; NPRS = NEPSY Response Set; NPVA = NEPSY Visual Attention; NPTW = NEPSY Tower; WJPC = WJ III COG Pair Cancellation; WJAWM = WJ IIICOG Auditory Working Memory; WJNR = WJ III COG Numbers Reversed; WJCF = WJ III COG Concept Formation; DKCW3 = D-KEFS Color Word Interference, Condition 3; DKTMT4 = D-KEFS Trail Making Test, Condition 4; DKVF3 = D-KEFS Verbal Fluency, Condition 3; DKCW4 = D-KEFS Color Word Interference, Condition 4; DKDF3 = D-KEFS Design Fluency, Condition 3; DKCS = D-KEFS Card Sorting; DK20Q = D-KEFS Twenty Questions; DKTOW = D-KEFS Tower; DKWC = D-KEFS Word Context; DKDF2 = D-KEFS Design Fluency, Condition 2.
* $p < .05$. Factor loadings are noted by the following subscripts: 1 = "Excellent"; 2 = "Very Good"; 3 = "Good"; 4 = "Fair"; 5 = "Poor"; 6 = "Uninterpretable."

Overall, there were similar patterns of strengths across all subsamples for the *Gsm* factor. The subscale scores of the subsample excluding individuals with learning disorders had relatively lower loadings on both the NEPSY Auditory Attention and NEPSY Response Set compared to the full sample and other subsamples. The second hypothesized factor, *Gs*, had a very high loading on the WJ III COG Pair Cancellation subtest (.65) and a poor loading on the D-KEFS Trail Making Test, Condition 4 subtest (.36) in the full sample. This pattern was consistent across all three subsamples.

The third hypothesized factor, *Gc*, had poor loadings on both the subscales (D-KEFS Word Context =.39, D-KEFS Twenty Questions=.37) in the full sample; however, this was inconsistent across the subsamples. In the subsample of boys only, compared to the full sample, D-KEFS Word Context loaded weaker (.27) and D-KEFS Twenty Questions loaded stronger (.45). Due to limitations of the analysis, path coefficients could not be calculated for the subsample excluding individuals with learning disabilities. This may be due to high variance in the performance on individuals with learning disabilities on the D-KEFS. For the subsample excluding individuals with ADHD, D-KEFS Word Context had an excellent path coefficient (.93), but D-KEFS Twenty Questions still loaded poorly. This discrepancy may be due to the low *N*'s for these subtests in each of the subsamples. For the fourth hypothesized factor, *Glr*, D-KEFS Color Word Interference, Condition 3 had a good path coefficient (.56), D-KEFS Verbal Fluency, Condition 3 loaded poorly, and D-KEFS Color Word Interference, Condition 4 had an excellent path coefficient (.88) in the full sample. This pattern was relatively consistent across the subsamples. On the fifth hypothesized factor, *Gf*, WJ III COG Concept Formation had a path coefficient that ranged from poor to good, with the weakest loading in the subsample of boys only (.38) and the strongest loading in the subsample excluding individuals with learning disabilities (.67). Across all subsamples, the path coefficient for the D-KEFS Sorting Test is not interpretable (<.32; Tabachnick & Fidell, 2001). For the final hypothesized model, *Gv*, the WJ III COG Planning and NEPSY Tower tests were negligible across the full sample and subsamples. The NEPSY Visual Attention subscale score had a poor factor loading for the subsample of boys only (.34) and was negligible across the full sample and other subscales (.15-.23). Across the full sample and all three subsamples, the D-KEFS Design Fluency Test, Condition 3 had a good path coefficient. D-KEFS Tower had negligible path coefficients across the full sample, the subsample of boys only, and the subsample that excluded individuals with learning disabilities (.06-.22); however, it had a good path coefficient in the subsample excluding individuals with ADHD (.76). The D-KEFS Design Fluency Test, Condition 2 had an excellent path coefficient in both the full sample and the subsamples excluding individuals with learning disabilities, a good path coefficient in the subsample of boys only, and a negligible path coefficient on the subsample excluding individuals with ADHD. This discrepancy may be due to the significant differences in means and standard deviations on this subtest, based on child's gender and disability category, as outlined in the MANOVA section.

As shown in Table 22, across the full sample and all subsamples the adjusted chi square values are all greater than 3, suggesting that the overall hypothesized model has a good fit with the data set of the current project, however the RMSEA and SRMR scores were all greater than .09, and the NNFI and CFI scores for the full sample and for all of the subsamples were all less than the critical level, further supporting the lack of fit of the overall model. Because of similarities in adjusted chi square values and fit indices, it cannot be concluded that model three is a better fit than both the single factor solution (e.g., model one) or Anderson et al.'s (2002) model solution (e.g., model two).

Model four (SNP model). **Is the underlying factor structure of executive functioning tasks across the WJ III COG, NEPSY, and D-KEFS in a clinical population of children, best described by the School Neuropsychological conceptual model (SNP model; Miller, 2007, 2010)?** The SNP model focuses specifically on the assessment of neuropsychological processes in children. Within the SNP model, executive functions are implicated as higher-order skills. According to the SNP model, components of executive functioning that are measurable by current neuropsychological assessments are as follows: Concept Generation, Inhibition, Motor Programming, Planning/ Reasoning/ Problem Solving, Set Shifting, Retrieval Fluency, Selective/ Focused Attention, Sustained Attention, the Use of Feedback in Task Performance, and Working Memory (Miller). However, due to the selection of subtests used in the current study, only the following constructs were analyzed: Inhibition, Planning/ Reasoning/ Problem Solving, Set Shifting, Selective/ Focused Attention, Sustained Attention, and Working Memory. The path diagrams with factor loadings on all subtests for this model are presented in Table 25 and Figure 6.

Table 25

Path Coefficients of Model Four Including the Full Sample and Three Sub-Samples

	Full Model	Males Only	No LD	No ADHD
Sustained Attention				
NPAA	.40 $*_5$	.35 $*_5$	.19 $*_6$	.38 $*_5$
NPVA	.28 $*_6$	.29 $*_6$	.16 $*_6$	.26 $*_6$
Selective Attention				
WJAA	.39 $*_5$	.33 $*_5$	.42 $*_5$	.38 $*_5$
WJPC	.65 $*_2$	.57 $*_3$	.70 $*_2$	.66 $*_2$
Working Memory				
WJAWM	.68 $*_2$	.66 $*_2$	.70 $*_2$	.65 $*_2$
WJNR	.72 $*_1$	.73 $*_1$	.71 $*_1$	.72 $*_1$
Inhibition				
NPRS	.50 $*_4$	.66 *	.42 $*_5$	.45 $*_4$
DKCW3	.36 $*_5$	.73 *	.41 $*_5$	.33 $*_5$
Problem Solving				
WJCF	.31 $*_6$	.27 $*_6$	.30 $*_6$	.10 $*_6$
WJPL	.66 $*_2$	.72 $*_1$	.76 $*_1$	.60 $*_3$
NPTW	.14 $*_6$	.16	.12	.29 $*_6$
DK20Q	.03	.15	.00	.11
DKTOW	.08	.00	.06	.05
DKWC	.19 $*_6$	.20 $*_6$	.19 $*_6$	.24 $*_6$
Set Shifting				
DKTMT4	.25 $*_6$	.16 $*_6$	.31 $*_6$	.20 $*_6$
DKVF3	.10	.11	.05	.09
DKCW4	.86 $*_1$	.91 $*_1$	.90 $*_1$	.92 $*_1$
DKDF3	.14	.20 $*_6$	.10	.10

Note. LD = Learning Disability; ADHD = Attention Deficit Hyperactivity Disorder; NPRS = NEPSY Response Set; NPVA = NEPSY Visual Attention; NPTW = NEPSY Tower; WJAA = WJ III COG Auditory Attention; WJPC = WJ III COG Pair Cancellation; WJAWM = WJ IIICOG Auditory Working Memory; WJNR = WJ III COG Numbers Reversed; WJCF = WJ III COG Concept Formation; DKCW3 = D-KEFS Color Word Interference, Condition 3; DKTMT4 = D-KEFS Trail Making Test, Condition 4; DKVF3 = D-KEFS Verbal Fluency, Condition 3; DKCW4 = D-KEFS Color Word Interference, Condition 4; DKDF3 = D-KEFS Design Fluency, Condition 3; DK20Q = D-KEFS Twenty Questions; DKTOW = D-KEFS Tower; DKWC = D-KEFS Word Context.
* $p < .05$. Factor loadings are noted by the following subscripts: 1 = "Excellent"; 2 = "Very Good"; 3 = "Good"; 4 = "Fair"; 5 = "Poor"; 6 = "Uninterpretable."

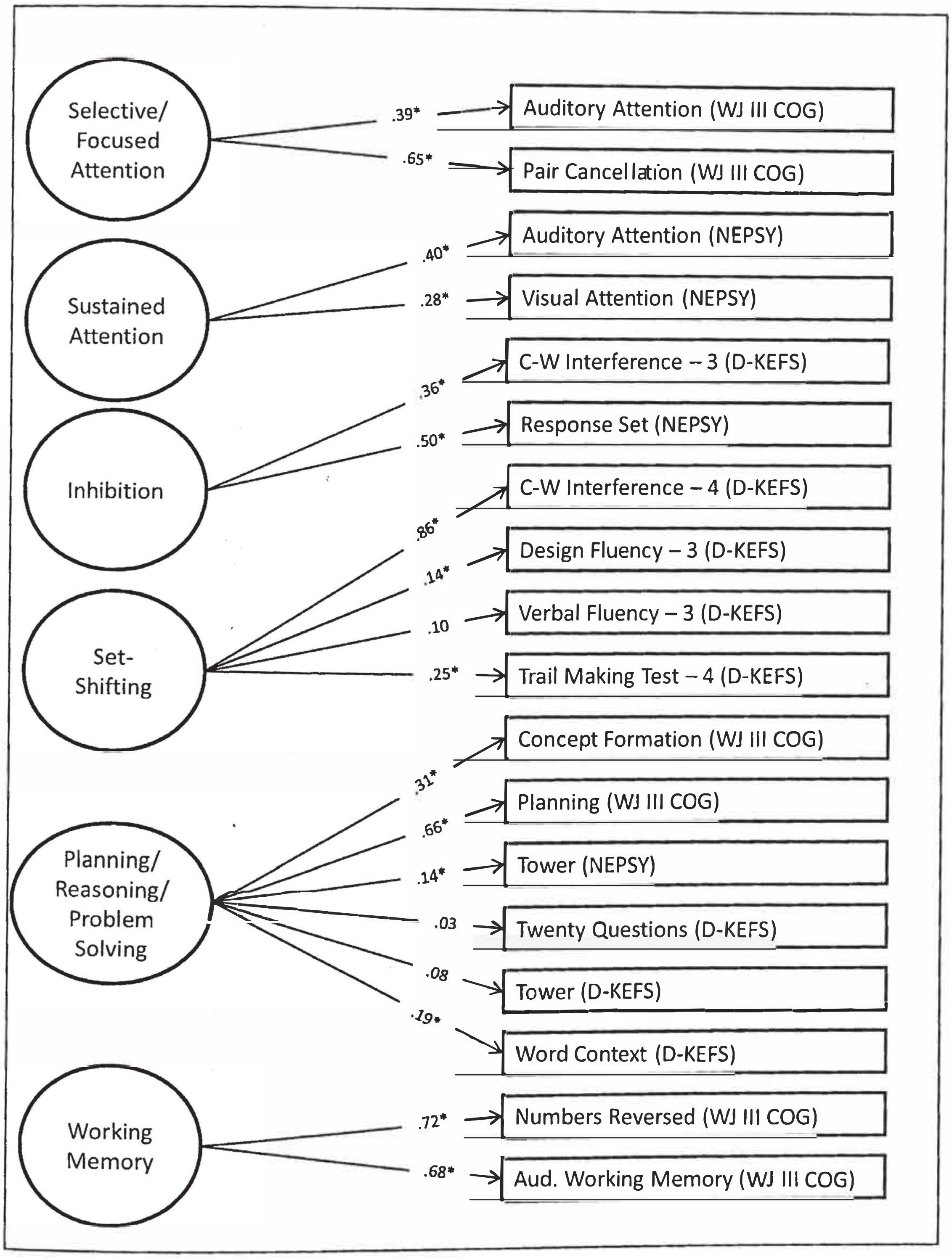

Figure 6. Illustration of Path Coefficients in Model 4 in the Full Sample

For the first factor, Sustained Attention, the NEPSY Auditory Attention subtest had poor path coefficients across the full sample and subsamples, with the exception of the subsample excluding individuals with learning disabilities, ranging from .35-.40 (.19 for no LD subsample). Across the full sample and all subsamples the NEPSY Visual Attention subtest did not have a meaningful path coefficient.

In regards to the second hypothesized factor, Selective Attention, WJ III COG Auditory Attention had a poor path coefficient across the full sample and all subsamples (.33-.42). Path coefficients for WJ III COG Pair Cancellation ranged from good (subsample of boys only, .57) to very good for the remaining samples (.65-.70). For the third hypothesized factor, Working Memory, path coefficients for the WJ III COG Auditory Working Memory subscale scores were very good and subscale scores for the WJ III COG Numbers Reversed had excellent path coefficients across the full sample and all of the subsamples. For the fourth hypothesized factor, Inhibition, the NEPSY Response Set subtest path coefficient was fair for the full sample (.50), very good for the boys only subsample (.66), and poor for the subsample excluding individuals with learning disorders and the subsample excluding individuals with ADHD (.42 & .45, respectively). For the full sample and the majority of the subsamples, D-KEFS Color Word Interference, Condition 3 had poor path coefficients (.33-.41); however in the boys only subsample there was an excellent path coefficient (.73). For the fifth hypothesized factor, Problem Solving, the WJ III COG Concept Formation, NEPSY Tower, and D-KEFS Twenty Questions, Tower and Word Context subtests had non-interpretable path coefficients across both the full sample and all of the subsamples (<.32). In the full sample and the sample excluding individuals with ADHD, WJ III COG Planning had very good path coefficients (.66 & .60, respectively). In both the boys only subsample and the subsample excluding individuals with learning disorders, the WJ III COG Planning had excellent path coefficients (.72 & .76, respectively). For the final hypothesized factor, Set Shifting, path coefficients for the D-KEFS Trail Making Test, Condition 4, D-KEFS Verbal Fluency, Condition 3 and D-KEFS Design Fluency Test, Condition 3 were not interpretable for the full sample or any of the subsamples (<.32). The D-KEFS Color Word Interference, Condition 4 had excellent path coefficients for the full sample and all of the subsamples, ranging from .86-.92.

As shown in Table 22, fit indices for the full sample indicate that the overall model is a good fit with the data collected (adjusted chi square = 3.185; RMSEA = .078; SRMR = .078; NNFI = .739; CFI = .796). Good RSMEA, SRMR, and CFI values were also found for the LD and ADHD removed subsamples. With the exception of the subsample removing individuals with learning disorders (adjusted chi square of 2.95), all adjusted chi square values were greater than 3, indicating a good fit. The good RMSEA, SRMR, and CFI fit indices for the LD removed subsample indicated that this model was a good fit for this subsample.

Based on the measures of fit, model four appears to be a much better fit with the data than models one, two, and three. Whereas the first three models appeared to fit in terms of the adjusted chi square values, the fourth model appeared to be a good fit based on adjusted chi square, CFI, RMSEA and SRMR, for the full sample and most of the subsamples. Furthermore, ECVI values were lower for most of the model four samples than the other three models, indicating that model four may be the most stable in the population. Given that model four had a greater statistical fit that the first three models, it can be concluded that given the current data set, this model appears to best explain the latent structure of executive functioning in the utilized sample of children.

Model four (SNP model-modified). **Is the underlying factor structure of executive functioning tasks across the WJ III COG, NEPSY, and D-KEFS in a clinical population of children best described by the School Neuropsychological (SNP) conceptual model?** In order to ensure the best model for the full sample, modifications were made to the hypothesized model. Given the nature of the design, there were three options in making such modifications. One option was to remove the set-shifting factor; however, conceptually this was not the best option as set-shifting is typically regarded as one of the primary aspects of executive functioning. Another option would be to use several unrelated factors as covariates (specifically, WJ III COG Pair Cancellations with Auditory Working Memory; NEPSY Tower with D-KEFS Twenty Questions; NEPSY Auditory Attention with NEPSY Response Set; D-KEFS Twenty Questions with Verbal Fluency, Condition 3; D-KEFS Twenty Questions with Color Word Interference, Condition 4; and D-KEFS Tower with Deign Fluency, Condition 3). Again, this would not make conceptual sense, because subtests that appeared to covary heavily statistically were not conceptually related. The final option was to add two covarying relationships (specifically, WJ III COG Pair Cancellations with Auditory Working Memory; and NEPSY Tower with D-KEFS Twenty Questions) to model 4. Given the research questions and data set, this option was considered to be optimal as it made the most conceptual sense, and was the most parsimonious option without disrupting the fit indices of the subsamples.

As shown in Table 26 and Figure 7, with the modifications to the model, the first factor, Sustained Attention, the path coefficients for the NEPSY Auditory Attention subscale in all subsamples (except the learning disorder removed sample) were poor (.36-.40). In the subsample without learning disorders, the path coefficient was not meaningful (.19). The path coefficients for the NEPSY Visual Attention subscale were also not meaningful for either the full sample or the subsamples (.16-.29). For the second factor, Selective Attention, path coefficients for the WJ III COG Auditory Attention were all poor, ranging from .35-.42. Path coefficients for WJ III COG Pair Cancellation were very good for the full sample, boys only subsample, and the ADHD removed subsample (.60-.70). Path coefficients were excellent for the learning disorder removed subsample (.72). For the Working Memory factor, path coefficients for the WJ III COG Auditory Working Memory subscale ranged from very good (full sample=.70; ADHD removed subsample=.68) to excellent (boys only subsample=.71; learning disorder removed subsample=.71). For the full sample and all of the subsamples, path coefficients for WJ III COG Numbers Reversed were all very good, ranging from .68 to .69. Path coefficients for NEPSY Response Set on the third factor, Inhibition, were fair for the full sample and all of the subsamples (.42-.51). Coefficients for D-KEFS Color Word Interference, Condition 3 were poor across all samples (.33-.41).

Table 26

Path Coefficients of Model Four with Modifications

	Full Model	Males Only	No LD	No ADHD
Sustained Attention				
NPAA	.40 $*_5$	.36 $*_5$	.19 $*_6$	.38 $*_5$
NPVA	.28 $*_6$	.29 $*_6$	.16 $*_6$	.26 $*_6$
Selective Attention				
WJAA	.39 $*_5$	.35 $*_5$	.42 $*_5$	.38 $*_5$
WJPC	.69 $*_2$	.60 $*_3$	.72 $*_1$	.70 $*_2$
Working Memory				
WJAWM	.70 $*_2$	.71 $*_1$	.71 $*_1$	.68 $*_2$
WJNR	.68 $*_2$	.68 $*_2$	.69 $*_2$	.68 $*_2$
Inhibition				
NPRS	.51 $*_4$	.45 $*_4$	.42 $*_5$	.45 $*_4$
DKCW3	.36 $*_5$	.38 $*_5$	.41 $*_5$	.33 $*_5$
Problem Solving				
NPTW	.32 $*_5$	.30 $*_6$	.30 $*_6$	.30 $*_6$
WJCF	.67 $*_2$	.71 $*_1$	.77 $*_1$	.63 $*_2$
WJGPL	.13	.14	.11	.09
DK20Q	.03	.18	.01	.10
DKTOW	.09	.00	.07	.06
DKWC	.19 $*_6$	.21 $*_6$	.19 $*_6$	.24 $*_6$
Set Shifting				
DKTMT4	.25 $*_6$	.15 $*_6$	.32 $*_5$	.20 $*_6$
DKVF3	.10	.11	.05	.09
DKCW4	.85 $*_1$	.97 $*_1$	.89 $*_1$	.96 $*_1$
DKDF3	.14 $*_6$	.20 $*_6$	.10	.10

Note. LD = Learning Disability; ADHD = Attention Deficit Hyperactivity Disorder; NPRS = NEPSY Response Set; NPVA = NEPSY Visual Attention; NPTW = NEPSY Tower; WJAA = WJ III COG Auditory Attention; WJPC = WJ III COG Pair Cancellation; WJAWM = WJ IIICOG Auditory Working Memory; WJNR = WJ III COG Numbers Reversed; WJCF = WJ III COG Concept Formation; DKCW3 = D-KEFS Color Word Interference, Condition 3; DKTMT4 = D-KEFS Trail Making Test, Condition 4; DKVF3 = D-KEFS Verbal Fluency, Condition 3; DKCW4 = D-KEFS Color Word Interference, Condition 4; DKDF3 = D-KEFS Design Fluency, Condition 3; DK20Q = D-KEFS Twenty Questions; DKTOW = D-KEFS Tower; DKWC = D-KEFS Word Context.
* $p < .05$. Factor loadings are noted by the following subscripts: 1 = "Excellent"; 2 = "Very Good"; 3 = "Good"; 4 = "Fair"; 5 = "Poor"; 6 = "Uninterpretable."

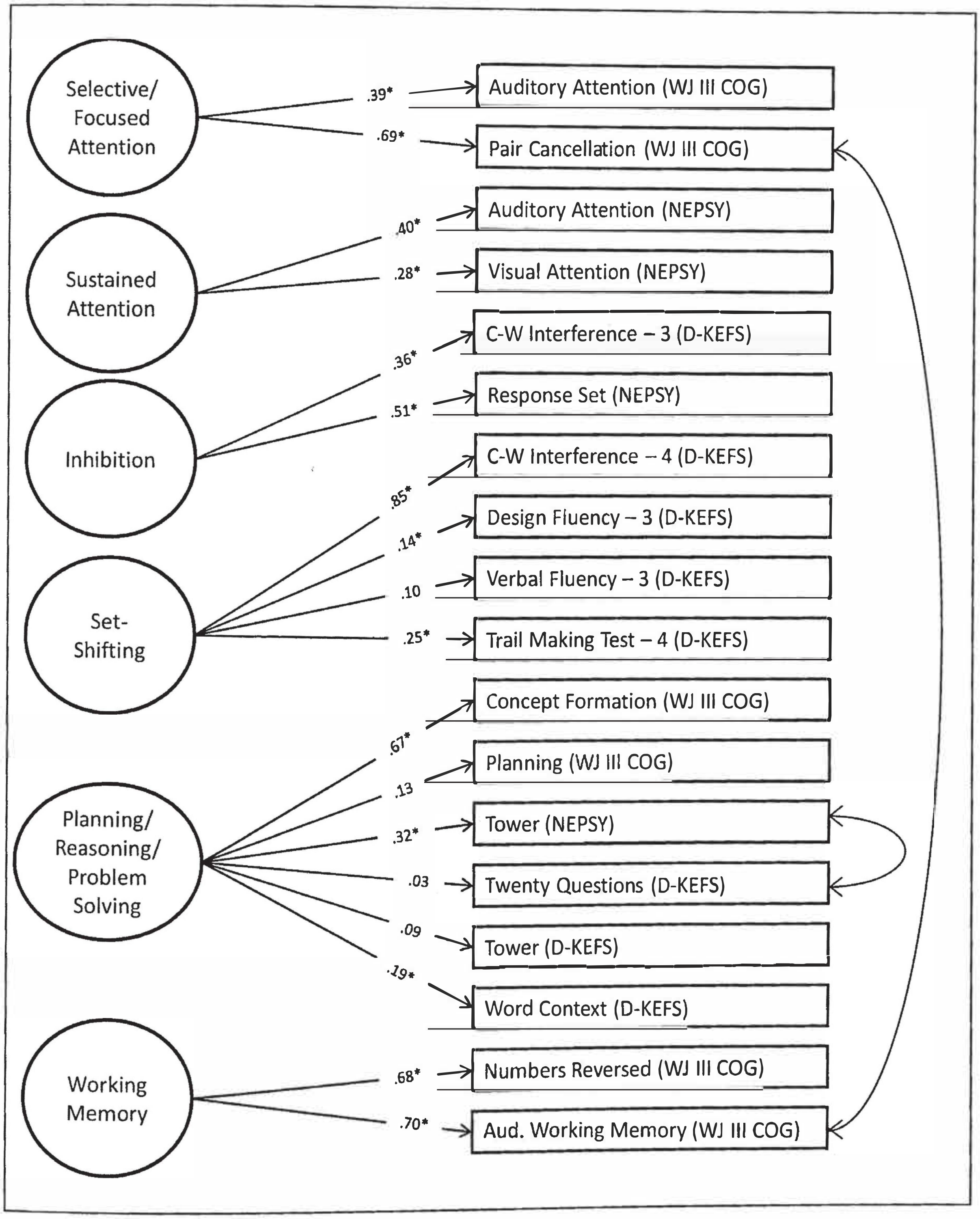

Figure 7. Illustration of Path Coefficients in Model 4 Modified in the Full Sample

On the fourth factor, Problem Solving, the WJ III COG Concept Formation had the highest path coefficients for the full sample and all subsamples, ranging from .63 to .77. Path coefficients for the NEPSY Tower, WJ III COG Planning, and D-KEFS Twenty Questions, Tower, and Word Context were not interpretable across the full sample or any subsamples, with the exception of the NEPSY Tower subtest in the full sample, which was poor (.32). For the final factor, Set Shifting, path coefficients for the D-KEFS Color Word Interference, Condition 4 test were excellent for the full sample and all subsamples (.85-.97). Path coefficients for all other subscales were not meaningful for the full sample or subsamples, with the exception of a poor path coefficient on the D-KEFS Trail Making Test, Condition 4 for the learning disorder removed subsample (.32).

As shown in Table 27, measures of fit for the fourth model with modifications indicate that the overall model is a good fit. With the exception of the learning disorder removed subsample, all adjusted chi square values were greater than 3 (LD removed subsample=2.925). RMSEA, SRMR, NNFI, CFI, and ECVI values further support the fit of the model with the full data set. Similar to the unmodified model four, with the modifications this model is still superior to models one, two, and three. Though indicators of fit are similar between both the unmodified and modified fourth model, path coefficients in the modified model are generally higher and more conclusive in the modified model, suggesting that the modified fourth model is the best statistical fit.

Table 27

Fit Indicies of Model Four with Modifications

	Full Sample	Boys Only	LD Removed	ADHD Removed
Adj Chi^2	3.074	3.318	2.925	3.493
RMSEA	.076	.094	.086	.089
SRMR	.076	.096	.085	.086
NNFI	.753	.592	.694	.670
CFI	.809	.686	.764	.745
ECVI	1.408	2.102	1.811	1.768

Note. Adj Chi^2 = adjusted chi-square; RMSEA = root mean square error of approximation; SRMR = standardized root mean squared residual; NNFI = Non-Normed Fit Index; CFI = comparative fit index; ECVI = single-sample cross validation index; LD = Learning Disability; ADHD = Attention Deficit Hyperactivity Disorder.

Chapter Summary

The purpose of this chapter was to present statistical results from the investigation of the underlying constructs of three neurocognitive tests in relation to four models of executive functioning. Executive functioning subtests within the WJ III COG, NEPSY, and D-KEFS were examined within: a broad, one factor model of executive functioning (model 1), Anderson and colleagues' (2002) model of executive functioning (model 2), a model depicting six of the broad CHC theory abilities (model 3) and a model depicting the SNP theory (model 4).

Initial descriptive statistics were first analyzed in order to obtain more information regarding the demographics and clinical make-up of the sample. Important categorical variables, such as gender, ethnicity, primary language, and disability category were compared to determine potential confounds in the study. Relationships among these demographic variables, as well as between the demographic variables and each subtest, were examined. Gender and disability criteria (specifically, LD and ADHD) significantly affected the subtest score means. Therefore, four different samples were utilized for each of the primary analyses (e.g., the full sample, sample with only males, sample with children with an ADHD diagnosis removed, and sample with children with a LD diagnosis removed).

Each of the four models were then examined with a series of confirmatory factor analyses. Results from the confirmatory factor analyses revealed that all models indicated significant adjusted chi-square values. However, goodness-of-fit indices indicated that the SNP theory model (model 4) offered significantly better fit to the data than the one factor model, the Anderson and colleagues' (2002) model, or the CHC theory model. The SNP theory model was then examined further to determine if any modifications, based on the significance and meaningfulness of each parameter, were needed. As a result of this process, two parameters were co-varied and confirmatory factor analyses were again performed on the modified SNP theory model (model 4-modified). The adjusted chi-square and goodness-of-fit indices for the modified model revealed a good overall fit to the sample data.

CHAPTER V

DISCUSSION

The current study examined underlying executive functioning constructs across the *Woodcock Johnson III Tests of Cognitive Abilities* (WJ III COG; McGrew et al., 2007; Woodcock et al., 2001c), the *Delis Kaplan Executive Function System (*D-KEFS; Delis et al., 2001), and *The NEPSY, A Developmental Neuropsychological Assessment* (NEPSY; Korkman et al., 1998). Four models of executive functioning were analyzed and compared to determine the best fit with the sample data. These models included a general, one-factor model of executive functioning, a factor model arising from Anderson and colleagues' (2002) theory of executive functioning, a model developed from the Cattell-Horn-Carroll theory of cognitive abilities (CHC theory; McGrew, 2005), and a model outlining executive functioning components from the Conceptual Model for School Neuropsychological Assessment (SNP model; Miller, 2007, 2010).

Purpose and Goals of Study

One of the goals of this study was to determine the validity of measuring executive functioning constructs within the WJ III COG, D-KEFS, and NEPSY, using concurrent validity methodology. These three measures can be utilized within neuropsychological test batteries; however, relatively limited concurrent and content validity evidence for these measures currently exists. According to test authors, the overall WJ III COG General Intellectual Ability (GIA) score has been compared to other overall intellectual scores (Woodcock et al., 2001c); however, little information on the validity of the Executive Processes, Working Memory, and Broad Attention cluster scales, when used with other executive functioning, working memory, or attention scales, has been noted. The NEPSY has also been evaluated and compared against other cognitive and neuropsychological scales for children. These studies identified that after comparisons with tests of similar domains, moderate correlations were found in the areas of attention, executive functioning, language, visual-spatial processing, and memory (Korkman et al., 1998). Evidence of validity in the D-KEFS was tested via correlations of the D-KEFS with several stand-alone measures of executive functioning, which resulted in preliminary convergent and discriminate validity evidence (Delis et al., 2001).

Although the WJ III COG, NEPSY, and D-KEFS test manuals report validity studies with other broad cognitive and/or stand alone executive functioning measures, only two studies have analyzed two of the three test batteries together. Carper (2003) examined correlations between subtests of the Executive Processes Cluster of the WJ III COG and the Tower and Design Fluency subtests of the NEPSY in a sample of typically developing children ($N = 60$). In addition, Floyd and colleagues (2006) identified relationships between the Executive Processes, Broad Attention, Working Memory, and Cognitive Fluency clinical clusters of the WJ III COG and the D-KEFS using correlation techniques in a sample of children ($N = 92$) and adults ($N = 100$). Although the aforementioned studies have contributed to the understanding of the clinical utility of these tasks, further validity evidence is needed. Thus, this study will add to the concurrent validity research base for the three identified measures.

Another goal of this study was to determine the fit of sample data to proposed theories of executive functioning. Although numerous theories that outline the structure and roles of the executive system have been described in the literature, none of these theories have yet to be validly incorporated into neuropsychological test batteries. Therefore, many executive functioning assessment tools that are available for use in children are either atheoretical or based on an unverified theory of executive functioning. This gap in the research has led to a lack of understanding of how executive functioning strengths or deficits actually manifest neurocognitively (Stuss & Alexander, 2000). Although a few studies have attempted to compare executive functioning tasks across test batteries, these studies are either based on a non-validated conceptual framework (Flanagan et al., 2010) or correlation based comparisons alone (Carter, 2003; Floyd et al., 2006). The use of confirmatory factor analytic techniques to determine the fit of executive functioning theories to specified executive functioning tests has been lacking in the literature. This is one of the first studies to evaluate the fit of data from neuropsychological assessments to various executive functioning theories.

The third purpose of this study was to determine the applicability of these three test batteries in a clinical population of children. Identifying the validity of these assessments in a sample of children will add to the knowledge of the nature of executive functioning development in children and the developmental pattern of frontal lobe functions. More specifically, measuring executive functioning tasks in a clinical group of children will further demonstrate if the selected subtests are good predictors of executive dysfunction. This knowledge could aid in distinguishing various childhood disorders often associated with frontal lobe dysfunction (Knight & Stuss, 2002) and may ensure more accurate clinical diagnoses, educational placements, and intervention strategies for children with executive functioning difficulties. However, research analyzing the validity of executive functioning tasks using a clinical sample of children is sparse in the literature. Therefore, one of the impetuses of this study was to provide information as to how various childhood clinical disorders may impact performance on executive functioning assessment tools.

Summary of Results

Preliminary Analysis

Preliminary analyses of the sample data were conducted to determine potential differences among the demographic categories utilized in the study. A significantly higher percentage of males than females were used in the study. This was expected due to the use of a clinical sample of participants and related research demonstrating the higher frequency of many school-aged clinical disorders in males (Gershon, 2002). Multivariate analyses revealed significantly different subtest performance, depending on gender, for several of the subtests utilized in the analysis. Of the significantly different subtests, females generally performed significantly higher than males, with the exception of the D-KEFS Design Fluency, Condition 2 test, on which males scored higher than females.

Although primary language and ethnicity were not reported in a large proportion of cases, the percentage of subjects that fell within each reported language and ethnicity category was similar to that of the standardization samples of the WJ III COG, NEPSY, and D-KEFS. Furthermore, as expected, the greatest percentage of English-speaking participants were Caucasian or African American and the greatest percentage of those who did not speak English were Hispanic or Latino American.

The proportion of different clinical diagnoses used in the study was not representative of frequencies of diagnoses in the general population. Rather, a higher frequency of neurologically-based diagnoses was analyzed, which was expected given the neuropsychological basis of the assessments administered to this sample. A significant relationship was found between language and clinical diagnosis. The majority of children with a learning disability (LD), autism spectrum disorder (ASD), or multiple disorders spoke English. However, a significantly higher proportion of participants with non-English primary languages were diagnosed with a language disorder. The high frequency of English Language Learners (ELL) diagnosed with a language disorder calls into question the accuracy or possible misdiagnosis of language disorders in ELL students, as it is difficult to distinguish a true language disability from linguistic and cultural differences in ELL students (Langdon & Wiig, 2009). Multivariate analyses revealed a significant effect of clinical diagnosis on subtest performance. This was especially true of individuals with an Attention Deficit Hyperactivity Disorder (ADHD) or a Learning Disability (LD) diagnosis.

Initial analyses of the sample's performance on WJ III COG, NEPSY, and D-KEFS subtests indicated that children with clinical diagnoses performed in the average range on WJ III COG subtests, and most NEPSY and D-KEFS subtests. The sample performed slightly below the average range on the NEPSY Response Set task and the D-KEFS Color Word Interference, Condition 3; Color Word Interference, Condition 4; Trailmaking, Condition 4; and Word Context tests. These tasks all require inhibition and/or mental flexibility. This may indicate that children with neurologically-based clinical disorders struggle more with the higher cognitive demands associated with these types of tasks. Interestingly, for most subtests the sample was restricted in range, with a low percentage of scores in the above average, high average, or below average range. The high percentage of low average scores could be anticipated, given the clinical nature of the sample. However, since subjects from differing clinical disorders and with assumed divergent cognitive abilities were included in the study, more variability among scores would be expected.

The results from the bivariate correlational analyses determined that all of the WJ III COG intercorrelations and most of the NEPSY intercorrelations reached statistical significance. The intercorrelations were consistent with the prior research discussed in the WJ III COG and NEPSY technical manuals (Korkman et al, 1998; McGrew & Woodcock, 2001; McGrew et al., 2007). However, nearly half (22 of 45) of the D-KEFS intercorrelations were not significant at the .01 or .05 level. This is comparable to the low test intercorrelations identified in the D-KEFS technical manual (Delis et al., 2001). The low intersubtest correlations may be explained by the atheoretical foundation of the D-KEFS. However, it is interesting that more D-KEFS tests did not correlate, especially since each of the tests were designed to measure an aspect of executive functioning. This suggests that many of the tests within the D-KEFS are measuring a distinct ability, rather than one, broad, executive functioning concept. This notion is also suggested in the D-KEFS technical manual, which states that the low correlations between D-KEFS tests "indicate that the instruments are not interchangeable and measure unique aspects of executive functioning" (Delis et al, pg. 82).

Correlations between test batteries were also analyzed in the current study. The WJ III COG Numbers Reversed and Auditory Working Memory subtests yielded the highest percentage of correlations (at the .01 level) with NEPSY subtests. Both of these subtests are included under the Working Memory cluster of the WJ III COG; thus, this may indicate that the NEPSY subtests include a working memory component. The NEPSY Tower task significantly correlated with every WJ III COG subtest, indicating high task comorbidity. Interestingly, this finding is dissimilar to Carper's (2003) study, which indicated that the NEPSY Tower task did not correlate with any of the subtests within the WJ III COG Executive Processes cluster. This difference may be due to the exclusion of children with clinical diagnoses in Carper's study, compared to the mixed, clinical sample utilized within the current study.

Similar to the intercorrelations between D-KEFS subtests, the majority of correlations between the D-KEFS and WJ III COG and NEPSY subtests were not significant. The WJ III COG Pair Cancellation and Concept Formation subtests resulted in the highest frequency of intercorrelations with the D-KEFS tests. Furthermore, the D-KEFS Color Word Interference, Condition 3 and Condition 4 tests resulted in a high frequency of intercorrelations in both tests batteries, as did the NEPSY Response Speed test. This suggests that these tasks are measuring a similar construct across batteries.

Of note, the WJ III COG Auditory Attention, D-KEFS Twenty Questions, and D-KEFS Tower tests each only significantly correlated with one other test from another battery. Thus, these tests may also be measuring a unique ability not shared with the other tests included in the study. The D-KEFS Tower test intercorrelations yielded results that were similar to the Floyd and colleagues (2006) study, which indicated that the D-KEFS Tower test did not correlate with any of the WJ III COG Executive Processes cluster subtests in a sample of children. Interestingly, the NEPSY Tower and D-KEFS Tower tests had a non-significant negative relationship. This was unexpected given the similar nature of each task. However, it should be noted that the Tower task was removed from the subsequent NEPSY II test battery, partially due to low concurrent validity evidence (Carper, 2003; Korkman et al, 2007b). This is further supported by contemporary research which suggests that not all tower tests can be used interchangeably (Baron, 2004; Goel et al., 2001; Humes et al., 1997).

Overall, much of the preliminary validity evidence presented in the current study was similar to previous research identified in the literature and in the WJ III COG, NEPSY, and D-KEFS technical manuals (Delis et al, 2001; Floyd et al., 2006; Korkman et al, 1998; McGrew & Woodcock, 2001; McGrew et al., 2007). However, some evidence was discrepant from prior research (Carper, 2003; Delis et al.; Floyd et al.; Korkman et al; McGrew & Woodcock; McGrew et al.), indicating that a clinical sample of children performs differently on executive functioning tests than samples of non-clinical children. Furthermore, within the clinical sample, gender and clinical diagnosis played a role in the outcome of test performance.

Primary Analysis

One of the primary goals of this study was to determine the model of executive functioning that best fits the factor structure of executive functioning tests among the WJ III COG, NEPSY, and D-KEFS in a clinical sample of children. Four *a priori* executive functioning models of increasing complexity were developed based on current research and literature. The four models selected included a general, one-factor model of executive functioning, a factor model arising from Anderson and colleagues' (2002) theory of executive functioning, a model developed from CHC theory (McGrew, 2005) and a model outlining executive functioning components of the SNP model (Miller, 2007, 2010). These models were analyzed using confirmatory factor analysis and compared to determine which model best described the sample data from the WJ III COG, NEPSY, and D-KEFS. Due to gender and diagnosis confounds identified in the multivariate analyses, each model was analyzed and compared using four different samples. The samples analyzed were: the full sample, a sample only including males, a sample with students with ADHD removed, and a sample with students with LD removed. Goodness-of-fit statistics were analyzed to determine adequate fit of sample data to each model. Furthermore, path coefficients were examined to determine the purest measures of factors. According to Comrey and Lee (1992) and Tabachnick and Fidell (2001), path coefficient loadings of .71 and greater are considered "excellent," loadings, between .63 and .70 are considered "very good," .55 to .62 are considered "good," .45 to .54 are considered "fair,".32 to .44 are considered "poor," and loadings less than .32 are "uninterpretable."

Research question one. Model one depicted a single factor solution, where all of the WJ III COG, NEPSY, and D-KEFS tests were specified as measures of one, broad executive functioning factor. This model was proposed as the "base" model to be compared with the more complex multifactor models. Since each of the test batteries were designed to measure more than one component of executive functioning, this model was predicted to be a poor representation of the sample data. As hypothesized, this model evidenced an inadequate fit to all four subsamples. Despite many of the path coefficients being interpretable, most fell below the "good" fit range. Although the adjusted chi-square values were adequate, poor fit was further confirmed by non-significant fit statistics. These findings are favorable, as it is generally desirable to rule out one-solution factors so that support for multi-factor solutions is stronger (Thompson, 2004).

Furthermore, these findings are consistent to results from contemporary literature. Model one was based on the theoretical view that all executive functioning tasks should load on one, broad factor (Anderson, 2008; Baddeley & Hitch, 1974; Blair, 2006; Duncan et al., 1996; Engle, 2002; Goldberg, 2002; Kane & Engle, 2002; Wiebe et al., 2011). Although this is a common viewpoint in the literature, it has received little direct research support. Specifically, except in a study that utilized a sample of 3-year-old children (Wiebe et al.) correlation studies have determined little to no association between executive functioning tasks and a broad, overarching cognitive ability (Friedman et al., 2006).

Anderson and colleagues' (2002) model of executive functioning (model two) was also examined to determine how well their three-factor model of executive functioning fit with the WJ III COG, NEPSY, and D-KEFS data. Due to the parsimonious, yet inclusive nature of this model, model two was hypothesized to be the best representation of the sample data. Although many interpretable path coefficients were evident, most were in the "poor" to "fair" range. Adjusted chi-square values were adequate across samples; however, fit indices did not indicate adequate fit to sample data: therefore, this model was not a good representation of sample data. Although parsimonious models are often ideal in confirmatory factor analysis, the poor fit of this model suggests that a more complex theory may best explain executive functioning within this sample.

The results of this analysis are not altogether surprising, due to the nature and limited research basis of this model. Like model one, model two was also based on a purely theoretical viewpoint that has yet to be substantiated with direct research (Anderson et al., 2002). Due to the theoretical nature of this model, associations between model factors and executive functioning subtests were never indicated by model authors. Therefore, task analyses and information provided by test manuals were utilized to determine the structure of model two. Since model two was based on a theory that has not yet taken assessment into account, the poor fit of the model to assessment data is understandable at this time.

Model three was developed to represent Stratum II of CHC theory (McGrew, 2005) using executive functioning subtests from the WJ III COG, NEPSY, and D-KEFS. Although seven of the CHC broad abilities were first proposed within this model, the *Ga* factor was removed, as the model could not be converged with the factor included. This was because only one subtest (WJ III COG Auditory Attention) could be conceptualized to load on the *Ga* factor, which is not ideal in confirmatory factor analysis (Tabachnick & Fidell, 2001). Thus, the WJ III COG Auditory Attention subtest was not included in this model. Similar to models one and two, model three yielded numerous "poor" and "uninterpretable" path coefficient loadings, as well as non-significant fit index values across all four samples. Because of similarities in adjusted chi square values and fit indices, it cannot be concluded that Model 3 is a better fit than either the single factor solution (model one) or Anderson and colleagues (2002) model solution (model two).

Although model three was based on the highly researched CHC theory (McGrew, 2005; McGrew et al., 2007), CHC theory does not describe a specific executive functioning factor. Rather, components of the executive system are integrated into broad and narrow ability factors (Flanagan et al., 2010; Kane & Engle, 2002; McGrew &

Woodcock, 2001). The WJ III COG is the only one of the three test batteries to directly apply CHC theory to individual subtests. Research has demonstrated that overall, the fit of the WJ III COG subtests and cluster scores to a CHC model is strong (McGrew et al.). Test authors have provided CFA and external validity support for the WJ III COG Working Memory clinical cluster (Shrank & Flanagan, 2003). However, although mentioned, validity support for the Executive Processes and Broad Attention clinical clusters are not described in detail in the technical manual (Shrank & Flanagan). This is consistent with the findings from this study which indicated that, with the exception of the WJ III COG Planning subtest, significant relationships between WJ III COG subtests and their associated CHC broad ability factor exist. Path coefficients for these subtests indicated "fair" (Concept Formation), "very good" (Auditory Working Memory and Pair Cancellation), and "excellent" (Numbers Reversed) loadings. The poor loading of the Planning subtest is consistent with the low reliability of that subtest, as reported in the WJ III COG technical manual (McGrew et al., 2007).

However, despite the support within the WJ III COG, limited research has been conducted to determine the fit of executive functioning subtests to the model across batteries. This is surprising, since CHC theory has served as the theoretical impetus to the development of the Cross-Battery Assessment Approach (XBA; Flanagan & McGrew, 1997; Flanagan et al., 2000). This lack of research support across batteries is consistent with the findings of this study. With the exception of the D-KEFS Color-Word Interference Condition 4 and Design Fluency, Condition 2 tests, subtests from the

NEPSY and D-KEFS loaded poorly on their respective CHC broad ability factors. This may indicate that the conceptual models proposed in the literature that have paired these subtests with CHC broad abilities may not be applicable in a clinical population of children (Flanagan et al., 2010).

Lastly, model four (SNP model; Miller, 2007, 2010) was developed to represent the fit of the executive functioning components of this model to the executive functioning subtests used in the current study. Several of the path coefficients across samples ranged from "excellent" to "good." However, much like models one through three, many path coefficients were in the "poor" to "fair" range or uninterpretable. However, unique to model four was the significance of various fit indices across samples, indicating that model four was a good fit to the sample data. Therefore, model four was a better fit with the data than models one, two or three. Whereas the first three models appeared to fit in terms of the adjusted chi square values, the fourth model appeared to be a good fit based on adjusted chi square, RMSEA, SRMR, NNFI, CFI, and ECVI values for the full sample and most of the subsamples. Therefore, it can be concluded that this model appears to best explain the latent structure of executive functioning in the utilized sample of children. Although it was not originally hypothesized that model four would represent the best fit to sample data, this finding is consistent with the theoretical foundations of each of the four models. Model four is the only model that was developed from a theory that focuses specifically on the assessment of neuropsychological processes in children (SNP model; Miller, 2007, 2010).

Final modified model. In order to potentially improve the overall factor loadings of model four, a modified *a posteriori* version of model four was developed. After analyzing the statistical structure of the model four factor analysis, as well as the theoretical meaningfulness of the data, a modified model was specified. Within this modified model, two covarying relationships (specifically, WJ III COG Pair Cancellation with Auditory Working Memory; and NEPSY Tower with D-KEFS Twenty Questions) were added to model four. Given the research questions and data set, this option was considered to be optimal as it made the most conceptual sense, and was the most parsimonious option without disrupting the fit indices of the subsamples. Measures of fit for the modified fourth model indicate that the modified model is a good fit across most samples. With the exception of the LD removed subsample, all adjusted chi square values were significant. Additionally, RMSEA, SRMR, NNFI, CFI, and ECVI values further support the fit of the model with the full data set. Similar to the unmodified model four, with the modifications this model is still superior to models one, two and three. Though indicators of fit are similar between both the unmodified and modified fourth model, path coefficients in the modified model are generally higher and more conclusive, suggesting that the modified fourth model represents the best statistical fit overall.

Although the LD removed subsample indicated lower overall fit, similar patterns among the path coefficients and goodness-of-fit statistics were evidenced across all four subsamples on the modified model four. A similar trend was noted on the model one, model two, model three, and unmodified model four analyses. Since patterns were similar across subsamples, the full sample group can be considered to be the most interpretable, due to the higher statistical power associated with a larger *n* (Comrey & Lee, 1992). However, it was interesting that the removal of the LD subsample of children resulted in more variability in the analyses. This may be due to the higher statistical variance within this subsample. This high variance may be expected in the LD subsample, as children with LD often demonstrate a discrepant pattern of cognitive strengths and weaknesses on neurocognitive tasks (Goldstein & Cunningham, 2009; Meltzer & Krishnan, 2007).

Although the modified fourth model represents the best statistical fit overall, some of the individual factors indicate stronger fit to the data than others. The factor with the strongest fit to sample data was the working memory factor. It should be noted that the two subtests included in the working memory factor are subtests from the WJ III COG Working Memory cluster. The strong fit of these subtests to sample data is consistent with contemporary literature that supports the grouping of tests within the WJ III COG Working Memory clinical cluster, evidenced by CFA analysis (McGrew et al., 2007; McGrew & Woodcock, 2001) and external validity studies with similar test batteries (Schrank & Flanagan, 2003; Wechsler, 1991). The factor with the second strongest factor loadings was the Selective Attention factor. Like the Working Memory factor, both of the subtests that were selected to load on this factor were WJ III COG subtests. Specifically, both of these subtests are described under the Broad Attention clinical cluster of the WJ III COG. Once again, it makes conceptual sense that these subtests loaded well together, as they are included in the same test battery. These findings were further substantiated by the validity evidence of these subtests and the Broad Attention cluster described within the *WJ III COG Normative Update Technical Manual* (McGrew et al.).

Of note, the Problem Solving factor, which included the most subtests across different batteries, had the poorest subtest loadings. This was a consistent pattern among many of the factors that utilized subtests from multiple batteries in models one through four. This impacts how cross-battery assessment should be used across these particular subtests in a clinical population of children. However, due to the limited scope and restricted population of this study, more research is needed to determine the utility of the XBA approach when assessing neurocognitive abilities. Several contemporary research articles and dissertations have supported the use of cross-battery assessment (Ganci, 2005; Hunt, 2008, Williams, 2005); however, the utility of this approach was not well represented in some assessments and populations (Morgan, 2008).

Within the modified model four, several individual subtests with reportedly low reliability coefficients (e.g., WJ III COG Planning; NEPSY Visual Attention; D-KEFS Twenty Questions; D-KEFS Tower; D-KEFS Design Fluency, Condition 3; D-KEFS Verbal Fluency, Condition 3; and D-KEFS Word Context), according to their respective test manuals (Delis et al., 2001; Korkman et al, 1998; McGrew & Woodcock, 2001; McGrew et al., 2007), also demonstrated weak factor loadings. A similar pattern was indicated across models and subsamples. This information, paired with the low reliability coefficients, indicates that these tasks may not be the best representations of executive functioning abilities in a clinical population of children. However, more research is needed to confirm this, as many of these tests are based on tasks that have historically been reputed to be measures of executive functioning, such as the trail-making test of the *Halstead-Reitan Neuropsychological Battery* (D'Amato & Hartlage, 2008; Halstead, 1952).

Furthermore, although the D-KEFS and NEPSY Tower tasks were based on historically significant tasks, such as the *Tower of London* (Shallice, 1982), *Tower of London-Drexel University* (Culbertson & Zillmer, 1998), or the *Tower of Hanoi* (Simon, 1975), their factor loadings were poor, as well as dissimilar, on the modified fourth model. This finding was similar across most models and subsamples. This is consistent with findings from contemporary research, which suggests that tower tasks should not be used interchangeably due to differences in the structure, neurocognitive demands, and cognitive abilities measured by each task (Baron, 2004; Goel et al., 2001; Humes et al., 1997). The results of the current study are further validated by research that indicates that tower tasks are not useful when assessing executive functioning in children, due to differences in the cognitive processes that are needed to complete the task, depending on the subjects age (Baker et al., 2001; Bishop et al. 2001).

Interestingly, the D-KEFS Color-Word Interference, Condition 3 test loaded poorly on the Inhibition factor, while the D-KEFS Color-Word Interference, Condition 4 test depicted an "excellent" loading on the Set-Shifting factor of the modified model four. This discrepancy is consistent with prior research on Stroop tasks, which has indicated that not all Stroop tasks are interchangeable and may not measure the same neurocognitive processes in children or adults (Salthouse & Meinz, 1995; Shilling et al., 2002).

Overall, the findings from the confirmatory factor analyses that were conducted in this study validate the use of the SNP model for assessing executive functioning skills in a clinical population of children. However, some of the subtests utilized in this study evidenced poor factor loadings, which were consistent with poor subtest reliabilities. This may indicate that these individual subtests are not actually the best representations of executive functioning and may need to be re-examined by test authors.

Implications to the Field of School Neuropsychology

Results from this study provided further validity evidence for the internal consistency of most of the WJ III COG and NEPSY subtests. However, internal consistency was not noted across the D-KEFS tests, further confirming that the D-KEFS tests are not interchangeable measures of executive functioning (Delis et al., 2001). A low frequency of significant correlations was also noted across the WJ III COG, NEPSY, and D-KEFS test batteries. This supports the theory that measures of executive functioning are actually assessing distinct skills, rather than a broad, universal concept (Ardila et al., 2000; Baron, 2004; McCabe et al., 2010). This idea was further supported by the results of the confirmatory factor analyses conducted in the current study, which suggest greater utility of the SNP conceptual model, as compared to a one-factor model, when assessing executive functioning skills in a clinical population of children.

The significant outcome of the SNP model confirmatory analysis, as well as the idea that tests of executive functioning are not interchangeable, should influence the use of cross-battery assessment with these tests. Cross-battery assessment techniques provide a systematic and valid way for practitioners to utilize multiple assessment tools for an evaluation, instead of being limited to only one instrument. More specifically, the concept of cross-battery assessment is also an essential component of neuropsychological assessments, as multiple instruments are typically needed in order to complete a comprehensive and valid neuropsychological evaluation (Flanagan & McGrew, 1997; Flanagan et al., 2010). Within the current sample, CHC theory was not a valid method for integrating executive functioning across batteries. However, these techniques may be still be applicable when utilized within a valid theoretical model, such as the SNP model.

Care should be taken when selecting subtests from different batteries, as each subtest appears to be measuring a distinct aspect of executive functioning. Some subtests may be more easily utilized across batteries than others, such as the subtests from the WJ III COG Working Memory clinical cluster, which evidenced significant correlations across batteries, as well as strong path coefficients within the SNP theoretical model. Furthermore, some subtests may not be the best representation of executive functioning when utilized within any model (e.g., WJ III COG Planning; NEPSY Visual Attention; D-KEFS Twenty Questions; D-KEFS Tower; D-KEFS Design Fluency, Condition 3; D-KEFS Verbal Fluency, Condition 3; and D-KEFS Word Context).

This study also validates the utility of these tests with children with clinical diagnoses. Although overall this sample's means were lower than those represented in each tests' standardization sample (Delis et al., 2001; Korkman et al, 1998; McGrew et al., 2007; McGrew & Woodcock, 2001), correlational analyses revealed the reliable use of most of these subtests within the clinical population. Low correlations were consistent with poor reliabilities noted in each tests' technical manual, indicating that the poor consistency may be due to problems within the individual subtests, rather than the use of a clinical population. This is an important finding in the field of school neuropsychology, as it can be assumed that the population of students that are given these assessments are students who are struggling academically or cognitively. Therefore, many of these children may have confirmed or unconfirmed clinical diagnoses. It is important for professionals to only use valid and reliable tests of executive functioning and to be aware of the different constructs that each executive functioning test is actually measuring. Continued research is also needed to further verify how children with executive dysfunction perform on these tasks.

Assumptions and Limitations

In order to analyze the results of this study, several assumptions were made. First of all, it was assumed that the executive functioning subtests of the WJ III COG, NEPSY, and D-KEFS were administered and scored in a standardized and valid format, consistent with the protocols outlined in each test manual. Consistency among testing procedures and techniques was expected because each of the professionals that contributed data were trained through the same program. Another assumption was that the clinical diagnoses assigned to subjects were accurate diagnoses based on consistent diagnostic criteria. However, this could be considered an area of variability in the current study, as the diagnosis of clinical disorders sometimes rests on the judgment of each practitioner.

Several limitations within the current study should also be noted. First of all, due to the use of archival data, the occurrence of missing data was present across the sample. However, this was accounted for by Multiple Imputation regression techniques; therefore the sample size and statistical power of this study were not significantly affected (Comrey & Lee, 1996; MacCollum et al., 1996). It should be noted that the large sample size may have affected the validity of some of the goodness-of-fit statistics utilized in the study (Barrett, 2007). However, this was accounted for by the concurrent use of multiple goodness-of-fit tests (Hu & Bentler, 1999).

Also of note, the sample was comprised of a variety of neuropsychological clinical groups, thereby creating a sample that was not representative of the general population of children as a whole. Consequently, consideration should be taken before applying the results of this study to other populations or circumstances. Another limitation was in the use of theories and subtests not extensively researched in the literature. This means that many of the results of the current study cannot be compared to prior research or literature; thereby making the validity of these results unclear. Therefore, results of this study should be interpreted with this consideration in mind.

Another limitation of the current study is in regards to one of the inherent weak points of confirmatory factor analysis. Whenever confirmatory factor analysis is used to test the fit of sample data to theoretical models, there is always a possibility that another, unspecified model is actually the best fitting example of that data. Therefore, conclusions that a particular model represents the best interpretation of sample data can never be made with definitive certainty.

Recommendations for Future Research

This study is the first to examine the executive functioning subtests of the WJ III COG, NEPSY, and D-KEFS in a clinical population of children. It is also the first to compare the relationships between this sample's performance on executive functioning tests to four theories of executive functioning. Therefore, this information offers a valid contribution to the field of pediatric and school neuropsychology. However, since this is the first study of its kind, replication of this study is essential to determining the validity of its results. Similar studies, utilizing exploratory and confirmatory factor analytic techniques would be beneficial in determining the applicability of other theories of executive functioning. Furthermore, extending this study to include recent neurocognitive batteries, such as the NEPSY II (Korkman et al., 2007a) would also provide additional and valuable validity in the realm of cross-battery assessment. Finally, it is recommended that this study be replicated in a general, non-clinical population of children, in order to provide additional information as to how executive dysfunction manifests in childhood.

Summary

In conclusion, the primary goals of the current study were to: further examine the concurrent validity of executive functioning tasks across the WJ III COG, NEPSY, and D-KEFS, determine the fit of these executive functioning tasks to four unsubstantiated theories, and predict the utility of using these tasks in a clinical population of children. Preliminary analyses revealed that gender and clinical diagnoses of subjects significantly influenced their performance on multiple subtests. Preliminary analyses also supported the internal consistency of the executive functioning subtests within the WJ III COG and NEPSY; however, this evidence was not supported within the D-KEFS. Intra-battery correlations and factor analyses indicated that individual executive functioning subtests appear to be measuring distinct abilities, rather than one broad, executive functioning factor. Therefore, not all of these subtests are interchangeable. Careful consideration should be used when utilizing cross-battery assessment techniques with these subtests. Some tests, such as the working memory subtests from the WJ III COG, may be more useful in cross-battery assessment than other tasks, such as the NEPSY or D-KEFS Tower tests.

Of particular note, the SNP model (Miller, 2007, 2010) indicated the best fit with sample data when compared to a one-factor executive functioning model, Anderson and colleagues (2002) model, and a model depicting the CHC theory of cognitive abilities (McGrew, 2005). This indicates that the SNP model describes the performance of children with clinical diagnosis on these subtests well. Although the clinical population of children utilized in this study performed differently than the standardization sample on these subtests, correlational analyses revealed the reliable use of most of these subtests within a clinical population. More research is needed to substantiate these findings and to generalize these results to other populations or test batteries.

REFERENCES

Ahmad, S. A., & Warriner, E. M. (2001). Review of the NEPSY: A developmental neuropsychological assessment. *The Clinical Neuropsychologist, 15*, 240-249.

Allison, P. D. (2003). Missing data techniques for structural equation modeling. *Journal of Abnormal Psychology, 112*, 545-557. doi:10.1037/0021-843X.112.4.545

Alvarez, J. A., & Emory, E. (2006). Executive function and the frontal lobes: A meta-analytic review. *Neuropsychology Review, 16*, 17-42. doi:10.1007/s11065-006-9002-x

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin, 103*(3), 411-423. doi:10.1037/0033-2909.103.3.411

Anderson, M. (2008). The concept and development of general intellectual ability. In J. Reed & J. Warner-Rogers (Eds.), *Child neuropsychology: Concepts, theory, and practice*. New York: Wiley-Blackwell.

Anderson, V. (1998). Assessing executive functions in children: Biological, psychological, and developmental considerations. *Neuropsychological rehabilitation, 8*, 319-349.

Anderson, V. (2002). Executive functions in children: Introduction. *Child Neuropsychology, 8*(2), 69-70.

Anderson, V., Levin, H. S., & Jacobs., R. (2002). Executive functioning after frontal lobe injury: A developmental perspective. In Stuss, D.T. & Knight, R.T. (eds.), *Principles of Frontal Lobe Function* (pp. 504-527). New York: Oxford University Press.

Ardila, A., Pineda, D., & Rosselli, M. (2000). Correlation between intelligence test scores and executive function measures. *Archives of Clinical Neuropsychology, 15*, 31-36. doi:10.1016/S0887-6177(98)00159-0

Arffa, S., Lovell, M., Podell, K., & Goldberg, E. (1998). Wisconsin Card Sorting Test performance in above average and superior school children: Relationship to intelligence and age. *Archives of Clinical Neuropsychology, 13*, 713-720. doi:10.1016/S0887-6177(98)00007-9

Army Individual Test Battery, (1944). Manual of directions and scoring, War Department, Adjutant General's Office, Washington, DC.

Avirett, E. K., & Maricle, D. E. (2011a). Executive dysfunction. In S. Goldstein, & J. A. Naglieri (Eds.), *Encyclopedia of child behavior and development*. Springer Science+Business Media LLC. doi:10.1007/978-0-387-77579-1_1046

Avirett, E., & Maricle, D. E. (2011b). Executive function. In S. Goldstein, & J. A. Naglieri (Eds.), *Encyclopedia of child behavior and development*. Springer Science+Business Media LLC. doi:10.1007/978-0-387-77579-1_1047

Baddeley, A. & Hitch, G. (1974). Working memory. In G.H. Bower (Ed.), *The psychology of learning and motivation* (pp. 47-89). New York: Academic Press.

Baddeley, A. D., & Hitch, G. J. (2000). Development of working memory: Should the Pascual-leone and the Baddeley and Hitch models be merged? *Journal of Experimental Child Psychology, 77*(2), 128-137. doi:10.1006/jecp.2000.2592

Baker, K., Segalowitz, S. J., & Ferlisi, M. (2001). The effect of differing scoring methods for the Tower of London task on developmental patterns of performance. *The Clinical Neuropsychologist, 15*, 309-313. doi:10.1076/clin.15.3.309.10273

Barkley, R. A. (2000). Commentary on the multimodal treatment study of children with ADHD. *Journal of Abnormal Child Psychology: An Official Publication of the International Society for Research in Child and Adolescent Psychopathology, 28*, 595-599. doi:10.1023/A:1005139300209

Baron, I. S. (2004). Delis-Kaplan Executive Function System. *Child Neuropsychology, 10*(2), 147-152.

Barrett, P. (2007). Structural equation modeling: Adjudging model fit. *Personality and Individual Differences, 42*, 815-824. doi:10.1016/j.paid.2006.09.018

Bayley, N. (1993). *Bayley Scales of Infant Development – Second Edition*. San Antonio, TX: Psychological Corporation.

Beaver, K. M., Wright, J. P., & Delisi, M. (2007). Self-control as an executive function: Reformulating Goddfredson and Hirchi's parental socialization thesis. *Criminal Justice and Behavior, 34*, 1345-1361.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*, 238-246. doi:10.1037/0033-2909.107.2.238

Benton, A. L., Hamsher, D., Varney, N. R., & Spreen, O. (1983). *Contributions to neuropsychological assessment: A clinical manual*. New York: Oxford University Press.

Bishop, D., Aamodt-Leeper, G., Creswell, C., McGurk, R., & Skuse, D. H. (2001). Individual differences in cognitive planning on the Tower of Hanoi task: Neuropsychological maturity or measurement error? *Journal of Child Psychology and Psychiatry, 42*, 551-556. doi:10.1111/1469-7610.00749

Blair, C. (2006). How similar are fluid cognition and general intelligence? A developmental neuroscience perspective on fluid cognition as an aspect of human cognitive ability. *Behavioral and Brain Sciences, 29*(2), 109-160.

Boll, T. (1993). *Children's Category Test.* San Antonio, TX: The Psychological Corporation.

Boone, K. B., Ponton, M. O., Gorsuch, R. L., González, J. J., & Miller, B. L. (1998). Factor analysis of four measures of prefrontal lobe functioning. *Archives of Clinical Neuropsychology, 13,* 585-595.

Bradshaw, J. L. (2001). *Developmental disorders of the frontostriatal system: Neuropsychological, neuropsychiatric, and evolutionary perspectives.* East Sussex: Psychology Press, LTD.

Brocki, K. C. & Bohlin (2004). Executive functions in children aged 6 to 13: A dimensional and developmental study. *Developmental Neuropsychology, 26*, 571-593.

Brookshire, B., Levin, H. S., Song, J. X., & Zhang, L. (2004). Components of executive function in typically developing and head-injured children. *Developmental Neuropsychology, 25*(1), 61-83.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park, CA: Sage.

Busch, R. M., McBride, A., Curtiss, G., Vanderploeg, D. (2005). The components of executive functioning in traumatic brain injury. *Journal of clinical and experimental neuropsychology, 27*, 1022-1032.

Byrne, B. M. (1989). *A primer of LISREL: Basic applications and programming for confirmatory factor analytic models.* New York: Springer-Verlag.

Carper, M. S. (2003). A concurrent validity study comparing executive functioning of the Woodcock Johnson III Tests of Cognitive Ability and the NEPSY (Unpublished Doctoral Dissertation). Texas Woman's University, Denton, TX.

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

Cohen, M. J. (1997). *Children's Memory Scale.* San Antonio, TX: The Psychological Corporation.

Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis (2nd ed.)*. Hillsdale, NJ England: Lawrence Erlbaum Associates, Inc.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory.* New York: Harcourt Brace Jovanovich College Publishers.

Culbertson, W. C., & Zillmer, E. A. (1998). The construct validity of the Tower of London-Drexel University as a measure of executive functioning of ADHD children. *Assessment, 5,* 215-226.

D'Amato, R. C., & Hartlage, L. C. (2008). In Hartlage L. C. (Ed.), *Essentials of neuropsychological assessment: Treatment planning for rehabilitation (2nd ed.)*. New York, NY US: Springer Publishing Co.

Dawson, P., & Guare, R. (2009). *Smart but scattered: The revolutionary "executive skills" approach to helping kids reach their potential.* The Guilford Press: New York.

Delis, D. C. (1988). *California Card Sorting Test*. Unpublished test.

Delis, D. C., Kaplan, E., & Kramer, J. (2001). *Delis Kaplan Executive Function System.* San Antonio, TX: The Psychological Corporation.

Delis, D. C., Kramer, J., Kaplan, E., & Ober, B. A. (2000). *California Verbal Learning Test – Second Edition. San* Antonia, TX: The Psychological Corporation.

Denckla, M. B. (1973). Development of speed in repetitive and successive finger-movements in normal children. *Developmental Medicine and Child Neurology, 15*, 635-645.

Donders, J. (1998). Cluster subtypes in the children's category test standardization sample. *Child Neuropsychology, 4*(3), 178-186. doi:10.1076/chin.4.3.178.3178

Duncan, J., Emslie, H., Williams, P., Johnson, R., & Freer, C. (1996). Intelligence and the frontal lobe: The organization of goal-directed behavior. *Cognitive Psychology, 30*(3), 257-303. doi:10.1006/cogp.1996.0008

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11*(1), 19-23. doi:10.1111/1467-8721.00160

Ewing-Cobbs, L., Prasad, M. R., Landry, S. H., Kramer, L., & DeLeon, R. (2004). Executive functions following traumatic brain injury in young children: A preliminary analysis. *Developmental Neuropsychology, 26*(1), 487-512.

Figner, B., Mackinlay, R. J., Wilkening, F., & Weber, E. U. (2009). Affective and deliberative processes in risky choice: Age differences in risk taking in the Columbia Card Task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 709-730. doi:10.1037/a0014983; 10.1037/a0014983

Flanagan, D. P., & McGrew, K. S. (1997). A cross-battery approach to assessing and interpreting cognitive abilities: Narrowing the gap between practice and cognitive science. In P. L. Harrison (Ed.), *Contemporary intellectual assessment: Theories, tests, and issues.* (pp. 314-325). New York, NY: Guilford Press.

Flanagan, D. P., Alfonso, V. C., Ortiz, S. O., & Dynda, A. M. (2010). Integrating cognitive assessment in school neuropsychological evaluations. In D. C. Miller (Ed.), *Best practices in school neuropsychology: Guidelines for effective practice, assessment, and evidence-based intervention* (pp. 101-140). Hoboken, NJ: John Wiley & Sons, Inc.

Flanagan, D. P., McGrew, K. S., & Ortiz, S. O. (2000). *The Wechsler intelligence scales and gf-gc theory: A contemporary approach to interpretation.* Needham Heights, MA: Allyn & Bacon.

Fletcher, J. M., Brookshire, B. L., Landry, S. H., & Bohan, T. P. (1996). Attentional skills and executive functions in children with early hydrocephalus. *Developmental Neuropsychology, 12*, 53-76. doi:10.1080/87565649609540640

Fletcher, J., & Taylor, H. (1984). Neuropsychological approaches to children: Towards a developmental neuropsychology. *Journal of Clinical and Experimental Neuropsychology, 6*, 39.

Floyd, R. G., McCormack, A. C., Ingram, E. L., Davis, A. E., Bergeron, R., & Hamilton, G. (2006). Relations between the Woodcock-Johnson III clinical clusters and measures of executive functions from the Delis-Kaplan Executive Function System. *Journal of Psychoeducational Assessment, 24*(4), 303-317. doi:10.1177/0734282906287823

Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., DeFries, J. C., & Hewitt, J. K. (2006). Not all executive functions are related to intelligence. *Psychological Science, 17*(2), 172-179. doi:10.1111/j.1467-9280.2006.01681.x

Fuster, J. M (2002). Physiology of executive functions: The perception-action cycle. In Stuss, D.T. & Knight, R.T. (eds.), *Principles of Frontal Lobe Function* (pp. 96-107). New York: Oxford University Press.

Ganci, M. (2005). The diagnostic validity of a Developmental Neuropsychological Assessment (NEPSY)-Wechsler Intelligence Scale for Children - Third Edition (WISC-III)-based cross-battery assessment. *Dissertation Abstracts International Section A*, 65.

Gershon J. (2002). A meta-analytic review of gender differences in ADHD. *Journal of Attentional Disorders*, *5*, 143-154.

Gioia, G., Isquith, P., & Guy, S. (2001). Assessment of executive functions in children with neurological impairment. *Psychological and Developmental Assessment: Children with Disabilities and Chronic Conditions,* 317.

Goel, V., Pullara, S. D., & Grafman, J. (2001). A computational model of frontal lobe dysfunction: Working memory and the Tower of Hanoi task. *Cognitive Science, 25*(2), 287.

Goldberg, E. (2002). *The executive brain: Frontal lobes and the civilized mind.* New York: Oxford University Press.

Golden, C., Freshwater, S. M., & Golden, Z. (2003). *Stroop Color and Word Test – Revised, children's version.* Wooddale, IL: Stoelting.

Goldstein, S., & Cunningham, S. (2009). Current issues in the assessment of intelligence, specific learning disability, and attention deficit hyperactivity disorder. In J. A. Naglieri, S. Goldstein (Eds.), *Practitioner's guide to assessing intelligence and achievement* (pp. 11-23). Hoboken, NJ US: John Wiley & Sons Inc.

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Erlbaum.

Gruber, S. A., Rogowska, J., Holcomb, P., Soraci, S., & Yurgelun-Todd, D. (2002). Stroop performance in normal control subjects: An fMRI study. *NeuroImage, 16*(2), 349. doi:10.1006/nimg.2002.1089

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

Halstead, W. (1952). The frontal lobes and the highest integrating capacities of man. *Halstead Papers*. M175, p. 26. Akron, OH: Archives of the History of American Psychology.

Heaton, R. (1981). *Wisconsin Card Sorting Test (WCST).* Odessa, FL: Psychological Assessment Resources.

Heaton, R. K., Chelune, G. J., Talley, J. L., Kay, G., & Curtiss, G. (1993). *Wisconsin Card Sorting Test.* Odessa, FL: Psychological Assessment Resources.

Homack, S., & Riccio, C. A. (2004). A meta-analysis of the sensitivity and specificity of the Stroop Color and Word Test with children. *Archives of Clinical Neuropsychology, 19*, 725-743. doi:10.1016/j.acn.2003.09.003

Homack, S., Lee, D., & Riccio, C. A. (2005). Test review: Delis-Kaplan Executive Function System. *Journal of Clinical and Experimental Neuropsychology, 27*, 599-609.

Hooper, S. R., Poon, K. K., Marcus, L., & Fine, C. (2006). Neuropsychological characteristics of school-age children with high-functioning autism: Performance on the NEPSY. *Child Neuropsychology, 12*, 299-305. doi:10.1080/09297040600737984

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

Horowitz, T., Schatz, B., & Schute, D. (1997). Trends in neuropsychological test usage. *Archives of Clinical Neuropsychology 12*, 338-339. doi:10.1093/arclin/12.4.338a

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1-55.

Hudspeth, W. J., & Pribram, K. H. (1990). Stages of brain and cognitive maturation. *Journal of Educational Psychology, 82*, 881-884.

Hughes, C. & Graham, A. (2002). Measuring executive functions in childhood: Problems and solutions. *Child and Adolescent Mental Health, 7*(3), 131-142.

Humes, G. E., Welsh, M. C., Retzlaff, P., & Cookson, N. (1997). Towers of Hanoi and London: Reliability of two executive function tasks. *Assessment, 4*(3), 249-257.

Hunt, M. S. (2008). A joint confirmatory factor analysis of the kaufman assessment battery for children, second edition, and the woodcock-johnson tests of cognitive abilities, third edition, with preschool children. *Dissertation Abstracts International Section A*, 68.

Jacobs, R., & Anderson, V. (2002). Planning and problem solving skills following focal frontal brain lesions in childhood: Analysis using the Tower of London. *Child Neuropsychology, 8*(2), 93-106. doi:10.1076/chin.8.2.93.8726

Joanes, D. N. & Gill, C. A. (1998). Comparing measures of sample skewness and kurtosis. *Journal of the Royal Statistical Society (Series D): The Statistician, 47*, 183–189. doi:10.1111/1467-9884.00122

Joseph, R. M., McGrath, L. M., & Tager-Flusberg, H. (2005). Executive dysfunction and its relation to language ability in verbal school-age children with autism. *Developmental Neuropsychology, 27*, 361-378. doi:10.1207/s15326942dn2703_4

Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic Bulletin & Review, 9*, 637-671.

Kaufman, A. S. & Kaufman, N. L. (2004). *Kaufman Assessment Battery for Children – Second Edition.* Circle Pines, MN: AGS Publishing.

Keith, T. Z. (2005). Using confirmatory factor analysis to aid in understanding the constructs measured by intelligence tests. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporaryintellectual assessment: Theories, tests, and issues* (2nd ed., pp. 581–614). New York:Guilford Press.

Kemp, S., Kirk, U., & Korkman, M. (2001). *Essentials of NEPSY Assessment*. New York: John Wiley & Sons, Inc.

Knight, R. T. & Stuss, D. T. (2002). Prefrontal cortex: The present and future. In Stuss, D.T. & Knight, R.T. (eds.), *Principles of Frontal Lobe Function* (pp. 573-597). New York: Oxford University Press.

Korkman, M. (1988). NEPSY – An adaptation of Luria's investigation for young children. *The Clinical Neuropsychologist, 2*, 375-292.

Korkman, M., Kirk, U., & Kemp, S. (1998). *NEPSY: A Developmental Neuropsychological Assessment*. San Antonio, TX: The Psychological Corporation.

Korkman, M., Kirk, U., & Kemp, S. (2007a). *NEPSY II: A developmental neuropsychological assessment: Administration Manual.* San Antonio, TX: The Psychological Corporation.

Korkman, M., Kirk, U., & Kemp, S. (2007b). *NEPSY II: A developmental neuropsychological assessment: Clinical and Interpretive Manual.* San Antonio, TX: The Psychological Corporation.

Langdon, H. W., & Wiig, E. H. (2009). Multicultural issues in test interpretation. *Seminars in Speech and Language*, 30(4), 261-278. doi:10.1055/s-0029-1241724

Latzman, R. D., & Markon, K. E. (2010). The factor structure and age-related factorial invariance of the Delis-Kaplan Executive Function System (D-KEFS). *Assessment*, 17(2), 172-184. doi:10.1177/1073191109356254

Levin, H. S., Culhane, K. A., Hartmann, J., Evankovich, K., Mattson, A. J., Harward, H., . . . & Fletcher, J. (1991). Developmental changes in performance on tests of purported frontal lobe functioning. *Developmental Neuropsychology, 7,* 377-395.

Lezak, M. D., Howieson, D. B., & Loring, D. W. (2004). *Neuropsychological assessment (4th ed.)*. New York, NY: Oxford University Press.

Luna, B., Garver, K. E., Urban, T. A., Lazar, N. A., & Sweeney, J. A. (2004). Maturation of cognitive processes from late childhood to adulthood. *Child Development, 75*, 1357-1372.

Luria, A.R. (1980). *Higher cortical functions in man* (2nd ed.). London: Penguin Press.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1*, 130-149. doi:10.1037/1082-989X.1.2.130

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods, 4,* 84-99.

Maricle, D. E., Johnson, W., & Avirett, E. (2010). Assessing and intervening in children with executive function disorders. In D. C. Miller (Ed.), *Best practices in school neuropsychology: Guidelines for effective practice, assessment, and evidence-based intervention* (pp. 599-640). Hoboken, NJ: John Wiley & Sons, Inc.

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

McCabe, D. P., Roediger, H. L., McDaniel, M. A., Balota, D. A., & Hambrick, D. Z. (2010). The relationship between working memory capacity and executive functioning: Evidence for a common executive attention construct. *Neuropsychology, 24*(2), 222-243. doi:10.1037/a0017619

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

McGrew, K. S., Schrank, F. A., & Woodcock, R. W. (2007). Technical Manual. *Woodcock Johnson III Normative Update.* Rolling Meadows, IL: Riverside Publishing.

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

Meltzer, L., & Krishnan, K. (2007). Executive function difficulties and learning disabilities: Understandings and misunderstandings. In Meltzer, L. (ed.), *Executive Function in Education* (77-105). New York: The Guilford Press.

Miller, D. C. (2007). *Essentials of School Neuropsychological Assessment.* Hoboken, NJ: John Wiley & Sons, Inc.

Miller, D.C. (2010). School Neuropsychological Assessment and Intervention. In D. C. Miller (Ed.), *Best practices in school neuropsychology: Guidelines for effective practice, assessment, and evidence-based intervention* (pp. 81-100). Hoboken, NJ: John Wiley & Sons, Inc.

Mirsky, A. F., Anthony, B. I., Duncan, C. C., Ahearn, M. B., & Kellan, S. G. (1991). Analysis of the elements of attention: A neuropsychological approach. *Neuropsychological Review, 2*(2), 109-145.

Mitchell, M. L., & Jolley, J. M. (2004). *Research Design Explained* (5^{th} ed.). Belmont, CA: Wadsworth, Thomson Learning Inc.

Morgan, K. (2008). The validity of intelligence tests using the Cattell-Hom-Carroll model of intelligence with a preschool population. *Dissertation Abstracts International Section A*, 69.

Mueller, R. O. (1997). Structural equation modeling: Back to basics. *Structural Equation Modeling, 4,* 353-369.

Nunnally J.C. (1967). *Psychometric Theory*. New York: McGraw-Hill.

Ozonoff, S., & Schetter, P. (2007). Executive dysfunction in autism spectrum disorders: From research to practice. In Meltzer, L. (ed.), *Executive Function in Education* (133-160). New York: The Guilford Press.

Parrish, J., Geary, E., Jones, J., Seth, R., Hermann, B., & Seidenberg, M. (2007). Executive functioning in childhood epilepsy: Parent-report and cognitive assessment. *Developmental Medicine & Child Neurology, 49*, 412-416. doi:10.1111/j.1469-8749.2007.00412.x

Pasini, A., Paloscia, C., Alessandrelli, R., Porfirio, M. C., & Curatolo, P. (2007). Attention and executive functions profile in drug naive ADHD subtypes. *Brain & Development, 29*(7), 400-408. doi:10.1016/j.braindev.2006.11.010

Pujol, J., Vendrell, P., Deus, J., Junque, C., Bello, J., Marti-Vilat, J. L., & Capdevila, A. (2001). The effect of medial frontal and posterior parietal demyelinating lesions on Stroop interference. *Neuroimage*, 13, 68-75.

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests (expanded edition)*. Chicago: The University of Chicago Press.

Reitan, R. M., & Wolfson, D. (1992). Conventional intelligence measurements and neuropsychological concepts of adaptive abilities. *Journal of Clinical Psychology, 48*, 521-529. doi:10.1002/1097-4679(199207)48:4

Riddle, R., Morton, A., Sampson, J. D., Vachha, B., & Adams, R. (2005). Performance on the NEPSY among children with spina bifida. *Archives of Clinical Neuropsychology, 20*, 243-248. doi:10.1016/j.acn.2004.07.004

Romine, C. B., Lee, D., Wolfe, M. E., Homack, S., George, C., & Riccio, C. A. (2004). Wisconsin Card Sorting Test with children: A meta-analytic study of sensitivity and specificity. *Archives of Clinical Neuropsychology, 19*, 1027-1041. doi:10.1016/j.acn.2003.12.009

Romine, C. B. & Reynolds, C. R. (2005). A model of the development of frontal lobe functioning: Findings from a meta-analysis. *Applied Neuropsychology, 12*, 190-201.

Rothbart, M. K., Sheese, B. E., & Posner, M. I. (2007). Executive attention and effortful control: Linking temperament, brain networks, and genes. *Child Development Perspectives, 1*(1), 2-7. doi:10.1111/j.1750-8606.2007.00002.x

Rubin, D. B. (1987) *Multiple Imputation for Nonresponse in Surveys*. New York, NY: John Wiley & Sons, Inc.

Salthouse, T. A., & Meinz, E. J. (1995). Aging, inhibition, working memory, and speed. *The Journals of Gerontology: Series B: Psychological Sciences and Social Sciences, 50B*(6), 297-306.

Schafer, J. L., & Olsen, M. K. (1998). Multiple imputation for multivariate missing-data problems: A data analyst's perspective. *Multivariate Behavioral Research, 33*, 545.

Schmitt, A. R., & Wodrich, D. L. (2004). Validation of a developmental neuropsychological assessment (NEPSY) through comparison of neurological, scholastic concerns, and control groups. *Archives of Clinical Neuropsychology, 19*, 1077-1093.

Schrank, F. A. (2005). Woodcock Johnson III Tests of Cognitive Abilities. In D.P. Flanagan & P.L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues.* New York: Guilford.

Schrank, F. A., & Flanagan, D. P. (2003). In Flanagan D. P. (Ed.), *WJ III clinical use and interpretation: Scientist-practitioner perspectives*. San Diego, CA: Academic Press.

Schrank, F. A., Flanagan, D. P., Woodcock, R. W., & Mascolo, J. T. (2002). *Essentials of WJ III cognitive abilities assessment*. Hoboken, NJ: John Wiley & Sons, Inc.

Schumacker, R. E., & Lomax, R. G. (2004). *A beginner's guide to structural equation modeling (2nd ed.)*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Shallice, T. (1982). Specific impairments of planning. In Broadbent & Weiskranz (Eds.), *Neuropsychological functions* (pp. 199-209). London: Royal Society.

Shallice, T., Marzocchi, G. M., Coser, S., Del Savio M., Meuter, R. F., & Rumiati, R. I. (2002). Executive function profile of children with attention deficit hyperactivity disorder. *Developmental Neuropsychology, 21*, 43-71.

Shilling, V. M., Chetwynd, A., & Rabbitt, P. M. (2002). Individual inconsistency across measures of inhibition: An investigation of the construct validity of inhibition in older adults. *Neuropsychologia, 40*, 605.

Simon, H. A. (1975). The functional equivalence of problem solving skills. *Cognitive Psychology, 7*, 268-288.

Slomine, B. S., Gerring, J. P., Grados, M. A., Vasa, R., Brady, K. D., Christensen, J. R., & Denckla, M. B. (2002). Performance on measures of 'executive function' following pediatric traumatic brain injury. *Brain Injury, 16*, 759-772. doi:10.1080/02699050210127286

Stein, J., & Krishnan, K. (2007). Nonverbal learning disabilities and executive function: The challenges of effective assessment and teaching. In Meltzer, L. (ed.), *Executive Function in Education* (106-132). New York: The Guilford Press.

Steinberg, L., Albert, D., Cauffman, E., Banich, M., Graham, S., & Woolard, J. (2008). Age differences in sensation seeking and impulsivity as indexed by behavior and self-report: Evidence for a dual systems model. *Developmental Psychology, 44*, 1764-1778. doi:10.1037/a0012955

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643-662. doi:10.1037/h0054651

Stuss, D. T., & Alexander, M. P. (2000). Executive functions and the frontal lobes: A conceptual view. *Psychological Research, 63*, 289-298.

Stuss, D. T., & Benson, D. F. (1986). *The frontal lobes.* New York: Raven Press.

Sylvester, C., Wager, T., Lacey, S., Hernandez, L., Nichols, T., & Smith, E. (2003). Switching attention and resolving interference: fMRI measures of executive functions. *Neuropsychologia, 41*, 357.

Szameitat, A. J., Schubert, T., Müller, K., & von Cramon, D. Y. (2002). Localization of executive functions in dual-task performance with fMRI. *Journal of Cognitive Neuroscience, 14*(8), 1184-1199. doi:10.1162/089892902760807195

Tabachnick, B. G., & Fidell, L. S. (2001). *Using Multivariate Statistics (4th ed.).* Needham Heights, MA: Allyn and Bacon.

Thompson, B. (2000). Ten commandments of structural equation modeling. In L. G. Grimm & P. R. Yarnold (Eds.), *Reading and understanding more multivariate statistics* (pp. 264-283). Washington, DC: American Psychological Association.

Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications.* Washington, DC: APA.

Thorndike, R. L., Hagen, E. P., & Sattler, J. S. (1986). *Stanford-Binet Intelligence Scale–Fourth Edition*. Itasca, IL: Riverside Publishing.

Vaquero, E., Gómez, C. M., Quintero, E. A., González-Rosa, J. J., & Márquez, J. (2008). Differential prefrontal-like deficit in children after cerebellar astrocytoma and medulloblastoma tumor. *Behavioral and Brain Functions, 4*, 1-16. doi:10.1186/1744-9081-4-18

Verté, S., Geurts, H., Roeyers, H., Oosterlaan, J., & Sergeant, J. A. (2006). Executive functioning in children with an autism spectrum disorder: Can we differentiate within the spectrum? *Journal of Autism & Developmental Disorders, 36*, 351-372. doi:10.1007/s10803-006-0074-5

Wang, L., Fan, X., & Willson, V. L. (1996). Effects of nonnormal data on parameter estimates and fit indices for a model with latent and manifest variables: An empirical study. *Structural Equation Modeling, 3*(3), 228-247. doi:10.1080/10705519609540042

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

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale – Third Edition*. San Antonio, TX: Psychological Corporation.

Welsh, M. C., Pennington, B. F., & Groisser, D. B. (1991). A normative-developmental study of executive function: A window of prefrontal function in children. *Developmental neuropsychology, 7(*2), 131-149.

Welsh, M. C., Pennington, B. F., Ozonoff, S., Rouse, B., & McCabe, E. R. (1990). Neuropsychology of early-treated phenylketonuria: Specific executive function deficits. *Child Development, 61*, 1697. doi:10.1111/1467-8624.ep9103040634

Wiebe, S. A., Sheffield, T., Nelson, J., Clark, C. C., Chevalier, N., & Espy, K. (2011). The structure of executive function in 3-year-olds. *Journal of Experimental Child Psychology*, 108(3), 436-452. doi:10.1016/j.jecp.2010.08.008

Willcutt, E. G., Doyle, A. E., Nigg, J. T., Faraone, S. V., & Pennington, B. F. (2005). Validity of the executive function theory of Attention-Deficit/Hyperactivity Disorder: A meta-analytic review. *Biological Psychiatry, 57*, 1336-1346. doi:10.1016/j.biopsych.2005.02.006

Williams, T. H. (2005). A joint-confirmatory factor analysis using the Woodcock-Johnson Tests of Cognitive Ability - Third Edition and the Stanford-Binet Intelligence Scales: Fifth Edition with high-achieving children. *Dissertation Abstracts International*, 66.

Wodka, E. L., Loftis, C., Mostofsky, S. H., Prahme, C., Larson, J. C. G., Denckla, M. B., & Mahone, E. M. (2008). Prediction of ADHD in boys and girls using the D-KEFS. *Archives of Clinical Neuropsychology, 23*, 283-293. doi:10.1016/j.acn.2007.12.004

Wood, J. N., & Grafman, J. (2003). Human prefrontal cortex: Processing and representational perspectives. *Nature Reviews Neuroscience, 4*(2), 139-147. doi:10.1038/nrn1033

Woodcock, R. W., & Johnson, M. B. (1977). *Woodcock-Johnson Psycho-Educational Battery.* Itasca, IL: Riverside Publishing.

Woodcock, R. W., & Johnson, M. B. (1989). *WJ-R Tests of Cognitive Ability.* Itasca, IL: Riverside Publishing.

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

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

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

Woodcock, R. W., McGrew, K. S., Mather, N., & Schrank, F. A. (2003). *Diagnostic Supplement to the Woodcock Johnson-III Tests of Cognitive Abilities.* Itasca, IL: Riverside Publishing.

Woodcock, R. W., Schrank, F. A., Mather, N., & McGrew, K. S. (2007). *Woodcock-Johnson III normative update*. Rolling Meadows, IL: Riverside Publishing.

Yadama, G. N., & Pandey, S. (1995). Effect of sample size on goodness-of-fit indices in structural equation models. *Journal of Social Service Research, 20*(3-4), 49-70. doi:10.1300/J079v20n03_03

Young, S. E., Friedman, N. P., Miyake, A., Willcutt, E. G., Corley, R. P., Haberstick, B. C., & Hewitt, J. K. (2009). Behavioral disinhibition: Liability for externalizing spectrum disorders and its genetic and environmental relation to response inhibition across adolescence. *Journal of Abnormal Psychology, 118*(1), 117-130. doi:10.1037/a0014657

Zelazo, P. D., Müller, U., Frye, D., & Marcovitch, S. (2003). The development of executive function. *Monographs of the Society for Research in Child Development, 68*(3), 1-27. doi:10.1111/j.0037-976X.2003.00261.x