COMPARING THE FACTOR STRUCTURES OF COGNITIVE MEASURES OF EXECUTIVE FUNCTION AND PARENT RATINGS OF EXECUTIVE FUNCTION IN A MIXED CLINICAL GROUP

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

MARISSA L. BENNERS, B.A.

DENTON, TEXAS

AUGUST 2017

ACKNOWLEDGMENTS

This project would not have been possible without the unwavering support and guidance of a number of individuals. First, I would like to thank my Chair and Advisor, Dr. Denise Maricle, who taught me foundational skills related to the field of school psychology, provided me with supervision and research opportunities, and has gone above and beyond to advocate for me. I would also like to thank Dr. Daniel C. Miller, who has shaped my understanding of school neuropsychology and also allowed me to access the KIDS, Inc dataset used as part of this study. Dr. DeOrnellas also played a key role in my education by equipping me with real-world, creative ideas to be applied when providing a variety of services. I also must thank my mentors, Amanda Gray and Ashley Fournier-Goodnight, as well as my cohort members, who provided me with support, motivation, recommendations to make it through each challenge encountered during my time in graduate school. This project also would not have been possible without my primary internship supervisor, Dr. Jordana Mortimer, who inspires me to work my hardest and to always surpass expectations and strive for my best.

Last, but certainly not least, I must thank my wonderful family members. My mother, Dr. Sherry Stewart, not only inspired me to pursue my PhD in the field of psychology, but also provided practical support by helping me edit papers and taking care of my children during classes, practicum, and internship. I also must thank my father, James Attrell, who is responsible for providing me with a personality characterized by unwavering focus and a drive to achieve, both of which led me to pursue and follow through with doctoral-level education. Next, I must thank my husband, Joshua Benners, who provided a resting place when I felt burnt out and encouraged me when I doubted myself. Without his constant love, support and patience, none of this would have been possible. Finally, I must thank my girls, Delilah Lynne and Eloise Jane, for all the joy they have brought to my life as I have worked to complete the final requirements of my doctoral program. I pray that my work will inspire you both to dream big and pursue those dreams.

ABSTRACT

MARISSA L. BENNERS

COMPARING THE FACTOR STRUCTURES OF COGNITIVE MEASURES OF EXECUTIVE FUNCTION AND PARENT RATINGS OF EXECUTIVE FUNCTION IN A MIXED CLINICAL GROUP

AUGUST 2017

The construct of executive function (EF) has been extensively researched in recent years due to its proposed role in a variety of clinical disorders. However, there remains contention between researchers regarding the definition and components of the construct (Eslinger, 1996; Packwood, Hodgetts, & Tremblay, 2011.) The current study sought to explore the construct of EF by investigating the factor structures of performance-based (i.e., cognitive) measures of EF as well as parent ratings of EF (i.e., behavioral measures). Data were culled from an archival database of neuropsychological case studies submitted as part of the KIDS, Inc School Neuropsychology Post-Graduate Certification Program. The database was narrowed using specific exclusion criteria. The final dataset consisted of 176 participants between the ages of 8 and 16. Two exploratory factor analyses were completed using full information maximum likelihood to account for missing data. The first analysis examined behavioral measures of EF and revealed a five factor structure. Factors were named Externalizing, Internalizing/Self-Regulation, Adaptive, Metacognition, and ADHD. The second analysis examined cognitive measures of EF and revealed a two factor structure. Factors were named Shifting and Reasoning.

Pearson product moment correlations were then computed to investigate the relationships between the factors obtained during each of the factor analyses. Correlations between behavioral and cognitive factors were weak. Lastly, five regression analyses (i.e., one for each of the five behavioral factors) were completed for each cognitive variable in order to determine whether the behavioral variables making up each factor predicted performance on cognitive variables. The Adaptive and Internalizing factors appeared to best predict performance on the D-KEFS variables, while the Adaptive, Internalizing, and Metacognition factors best predicted performance on the NEPSY-2 variables. Overall, the factors were poor predictors of performance on the WJ III COG NU. The results of this study provide additional evidence that the construct of EF is highly complex and thus difficult to define and effectively measure. Furthermore, behavioral and cognitive measures of EF should not be considered interchangeable; instead, they should be considered separate but key components of neuropsychological assessment.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

INTRODUCTION

Executive function (EF) can be defined as a set of higher-order abilities, such as attention, planning, organization, inhibition, working memory, and problem-solving, that facilitate goal-directed thought and behavior (Hunter & Sparrow, 2012; McCloskey, Perkins, & Van Divner, 2009; Packwood, Hodgetts, & Tremblay, 2011). Lack of adequate EF abilities may result in executive dysfunction, which can contribute to difficulty completing everyday activities and engaging with others in a socially appropriate manner. Multiple definitions and theories of EF have been outlined by researchers but no single agreed upon theory of the concept currently exists. This is due to the highly complex nature of EF. Higher-order abilities rely strongly on more primitive skills, making it difficult to create a pure measure of EF.

Both unitary and multimodal theories of EF exist (Flanagan & Harrison, 2012). Unitary theories view EF as one ability that oversees lower-order abilities, whereas multimodal theories view EF as an umbrella term that includes various, but separate, cognitive processes. Most recent theories view EF as multimodal. Some of the most popular theories of EF include Baddeley and Hitch's Model of Working Memory (2000), Denckla's theory of EF (1996), Lezak's theory of EF (1995, 2004), Anderson's Executive Control System Model (2002), the Supervisory Activating System Model (Norman & Shallice, 1986), Barkley's self-regulatory model (1997), hot and cool EF

(Abelson, 1963), and McCloskey's theory of EF (McCloskey et al., 2009). Other theories and models recognize EF as a key component of intellectual and neuropsychological functioning. These theories include the Cattell-Horn-Carroll (CHC) and Integrated School Neuropsychology/CHC models (Flanagan & Harrison, 2012; Miller, 2013). These theories view and define EF in very different ways but each recognizes the importance of EF as a higher-order ability in a variety of sit uations and settings. A recent meta-analysis showed that 68 terms have been used to describe EF and over 98 tasks have been developed to assess EF based on the above theories and others (Packwood et al., 2011).

Abridged Literature Review

History of Executive Function

The brain was first identified as a source of mental processes by the famous philosopher, Plato (Simon, 1972). Other key figures, including Rene Descartes, Thomas Willis, Franz Josef Gall, Paul Broca, and Carl Wernicke contributed to the idea that localized areas of the brain contribute to specific processes (Kaitaro, 2001; Schultz & Schultz, 2008). The railroad accident of Phineas Gage (Harlow, 1848) contributed significantly to today's understanding of the role of the brain in EF abilities. After experiencing left frontal lobe damage when a large iron rod passed through his brain, Phineas Gage exhibited extreme behavioral and personality changes, including difficulty regulating his emotions. This demonstrated the importance of the frontal lobe in self-regulation and inhibition (Aron, 2008; Barkley, 2012b).

Alexander Luria (1966) was the first to introduce the concept of EF as it is known today by proposing that damage to the prefrontal cortex (PFC) results in disruption of the brain's critical faculty, which is responsible for overseeing behavior (Hunter & Sparrow, 2012). He coined the term *frontal lobe syndrome* to define deficits resulting from frontal lobe impairment. After Karl Pribham (1973) used the term *executive* to describe the functions of the PFC, *frontal lobe syndrome* was renamed *executive disorder* and then *dysexecutive syndrome* (Baddeley, 1986; Fuster, 1997).

Neurobiology of Executive Function

The PFC is the primary area of the brain associated with EF abilities (McCloskey et al., 2009; Poletti, 2009). Damage to the frontal region often results in poor performance on measures of EF (Aron, 2008; Blumenfeld, 2010; Goldberg, 2001). Specific areas within the PFC appear to underlie certain EF skills (Sonuga-Barke, 2002; Zelazo & Carlson, 2012). The dorsolateral PFC is thought to play a role in cognitive aspects of EF while the orbitofrontal PFC and its connections to subcortical regions of the brain are thought to play a role in emotional aspects of EF. The anterior cingulate cortex is particularly important in regards to motivation and goal-directed decision-making (Shallice, Marzocchi, Coser, del Savio, Meuter, & Rumiati, 2002). Cascade theory posits that the various connections between regions of the PFC and other areas of the brain facilitate EF (Banich, 2009). In other words, the PFC is not the only region responsible for EF; instead, complex neural networks throughout the brain determine self-regulation.

Multiple neurotransmitters play a role in overall EF ability (Logue & Gould, 2014). These neurotransmitters include dopamine, serotonin, norepinephrine, and acetylcholine. Because the PFC is made of pyramidal cells, it is very susceptible to even small disruptions in neurotransmitter levels (Hosenbocus & Chahal, 2012). Low levels of dopamine and norepinephrine in the PFC have been associated with deficits in attention and set-shifting while low levels of serotonin have been linked to poor inhibition (Logue & Gould, 2014). Increases in these neurotransmitters may result in enhanced EF abilities. The cholinergic system may mediate these changes by influencing dopamine levels.

Development of Executive Function

Research consistently supports the notion that EF skills begin to develop early in life and become increasingly complex over time (Zelazo, Muller, Frye, & Marcovitch, 2003). During infancy, humans develop the abilities to attend to the world around them and engage in very basic problem-solving (Hunter & Sparrow, 2012). A key period of EF development occurs during toddlerhood when children begin to develop working memory, set shifting, and inhibition skills. Cognitive flexibility, planning, and decision-making then arise during middle childhood. Adolescence is another critical period of EF development as multiple EF abilities become more complex. However, behavioral components of EF, particularly behavioral inhibition, do not mature until early adulthood. EF abilities often peak during this stage and gradually decline with age (Banich, 2009). The trajectory of EF development coincides with physical growth spurts and neural proliferation in the PFC (Jurado & Rosselli, 2007). Critical periods of EF development coincide with periods of brain plasticity during which the brain is highly susceptible to change through environmental experiences and stimulation (Zelazo & Carlson, 2012).

Multiple factors contribute to one's ability to successfully engage in EF-related tasks. Genetics may explain up to 99% of certain EF skills, including inhibition and shifting (Friedman, Miyake, Young, Defries, Corley, & Hewitt, 2008). Genes linked to

EF include catechol-amethyltransferase (COMT) and reelin genes (Baune et al., 2010; Logue & Gould, 2014). Environmental factors influencing EF include maternal depression during pregnancy and exposure to teratogens, such as drugs, alcohol, lead, mercury, and pesticides (Dawson & Guare, 2010; Hughes, Roman, Hart, & Ensor, 2012; Riccio, Sullivan, & Cohen, 2010). Additionally, harsh rearing, neglectful parenting, and low cognitive stimulation during childhood may contribute to poor EF skills throughout the lifespan (Rhoades, Greenberg, Lanza, & Blair, 2011). Finally, certain cultural factors appear to affect EF ability. Most notably, socioeconomic status (SES) has been found to consistently predict EF by altering neural development through stress, infection, poor nutrition, and other factors (Noble, McCandliss, & Farrah, 2007). It is likely that the above environmental factors influence EF by turning genes that underlie EF skills on or off through epigenetic changes (Champagne, 2010).

Clinical Diagnoses and Executive Function

Children diagnosed with neurodevelopmental and mental health disorders, including attention-deficit/hyperactivity disorder (ADHD), autism spectrum disorder, traumatic brain injury (TBI), learning disabilities, depression, and anxiety often exhibit unique patterns of EF deficits (Pennington & Ozonoff, 1996). ADHD commonly involves poor attentional control and behavioral inhibition and some theorists characterize ADHD as a disorder of EF itself (Barkley, 1997). EF deficits often apparent in children with autism include poor cognitive flexibility, planning, organization, initiation, and working memory skills (Gilotty, Kenworthy, Sirian, Black, & Wagner, 2002; Ozonoff & Cathcart, 1998). A variety of EF deficits may arise following TBI depending upon the severity and location of the injury (Gioia, Isquith, Kenworthy, & Barton, 2002). Learning disabilities may involve deficits in working memory, problem-solving, and motivation. Finally, depression has been associated with poor self-regulation, problem-solving, and initiation, while anxiety has been associated with deficits in cognitive shifting and inhibition (Hosenbocus & Chahal, 2012; Hunter & Sparrow, 2012).

Assessment of Executive Function

The construct of EF has proven difficult to assess due to lack of clarity regarding the definition of EF, the use of time-limited performance-based testing, artificial test environments, and poor ecological validity (Gioia et al., 2002). Test instruments often do not provide a full picture of a child's EF abilities and scores may be artificially inflated due to test environments that facilitate attention and reduce demands on problem-solving. Thus, performance on cognitive measures of EF may not generalize to natural settings (Slick, Lautzenhiser, Sherman, & Eyrl, 2006; Silver, 2000). Classic measures of EF include the Wisconsin Cart Sorting Test (WCST; Heaton, Chelune, Talley, Kay, & Curiss, 1993), Stroop Color and Word Test (Golden & Freshwater, 2002), trail-making tests, go-no-go tests, and tower tests (Davis, 2011). These tasks inspired the development of subtests included in test batteries that were used as part of the current study, such as the Delis Kaplan Executive Function System (D-KEFS; Delis, Kaplan, & Kramer, 2001), NEPSY-2 (Korkman, Kirk, & Kemp, 2007), Test of Everyday Attention for Children (TEA-Ch; Manly, Robertson, Anderson, & Nimmo-Smith, 1999), and Woodcock Johnson Tests of Cognitive Abilities, Third Edition (WJ III COG NU; McGrew, Schrank, & Woodcock, 2007).

Behavioral rating scales were created to address some of the limitations of cognitive measures of EF. Rating scales are known to have higher levels of validity because they more accurately predict behaviors in real-world settings (Franzen & Wilhelm, 1996; McCloskey et al., 2009). However, they are limited by rater bias, which occurs when raters under- or over-report behaviors. Behavioral rating scales measuring EF include the Behavior Rating Inventory of Executive Function (BRIEF; Gioia Isquith, Guy, & Kenworthy, 2000), Barkley Deficits in Executive Functioning Scale: Children and Adolescents (BDEFS: CA, Barkley, 2012a), and Delis-Rating of Executive Functions (D-REF; Delis, 2012). The Behavior Assessment System for Children, Second Edition (BASC-2; Reynolds & Kamphaus, 2004) also provides information regarding patterns of behavior that may be indicative of poor EF abilities.

Rationale, Significance, and Purpose of the Current Study

The primary purpose of this study was to determine the relationship between cognitive (i.e., performance-based) measures of EF and behavioral ratings of EF in a mixed clinical sample by answering three primary research questions. These questions involved the comparison of scores obtained on cognitive measures of EF and parent ratings of behavioral EF. The questions were as follows:

1. What factor structures are obtained for both behavioral EF (i.e., parent ratings) and cognitive EF (i.e. performance-based measures)?
2. Are the factor structures obtained for behavioral EF and cognitive EF comparable? In other words, do the behavioral EF factors correlate with the cognitive EF factors?

3. Finally, do the subtests loading on behavioral EF factors predict those loading on correlated cognitive EF factors?

It was hypothesized that multiple factors would be obtained for both the behavioral ratings of EF and the cognitive measures of EF. Specifically, behavioral measures were expected to load onto two separate factors as shown during factor analyses examining the BRIEF (Gioia et al., 2000). Cognitive measures were expected to load onto multiple factors. This hypothesis was based on the Integrated SNP/CHC model (Miller, 2013), which identifies four domains of EF along with related constructs included in the domain of Cognitive Facilitators/Inhibitors. Regarding the second and third research questions, it was anticipated that there would be some degree of similarity between the behavioral and cognitive factors extracted and that the factors would be so similar that behavioral measures would predict related cognitive measures. These hypotheses were based upon research demonstrating strong correlations between the BRIEF, BASC-2, and classical measures of EF (Reynolds & Kamphaus, 2004) and predictive relationships between ratings on the Brown ADD Scales (Brown, 2001) and performance on the NEPSY-2 (Korkman et al., 2007). That being said, it was expected that there would be some degree of difference between the factors obtained on the two EFAs given research revealing low correlations between scores on EF tasks and behavioral ratings by caregivers (Wilson, 1998). This study was designed to help clarify the degree to which cognitive and behavioral measures of EF can be compared.

Significance of the Current Study

The study of EF has accelerated in recent years, leading to a variety of definitions and theories of the construct; however, no single definition or theory has been widely accepted. Because EF has not been well-defined, it has led to disagreement regarding the most valid means of assessing EF. The current study was designed to determine the relationship between scores on cognitive measures of EF and parent ratings of EF in a mixed clinical sample. Results were expected to contribute to the field of psychology by providing information regarding the validity of EF behavioral rating scales, which have become more readily available in recent years. If cognitive and behavioral measures of EF were found to be strongly related and predictive of one another, EF scales may be considered a useful alternative to cognitive testing in screening children for EF deficits.

CHAPTER II

LITERATURE REVIEW

This chapter will provide information relevant to the current study, including popular definitions and theories of executive function (EF), the history of EF, the neurobiology of EF, the developmental trajectory of EF, the role of EF in various clinical disorders, and the assessment of EF through cognitive and behavioral measures. This foundational knowledge will serve as a rationale for completing the current study.

Definition and Theories of Executive Function

The concept of executive function (EF) has been researched extensively in recent years due to its proposed role in a variety of clinical disorders, such as ADHD and autism (McCloskey et al., 2009). The definition of EF has been contested by researchers over the years (Eslinger, 1996). Packwood et al. (2011) conducted a meta-analysis, which showed that nearly 68 terms are used to define components of EF and over 98 tasks have been developed to test these components. The skills most commonly associated with EF include attention (i.e., selective, shifting, and attentional control), planning, cognitive and behavioral inhibition, working memory, and self-monitoring. Due to great variation between the definitions of EF, a testable and concise theory of the concept does not yet exist. Although no single definition has been identified, most researchers agree that EF is comprised of higher-order abilities that facilitate goal-directed thoughts and behavior through attentional control, behavior regulation, planning, organization, and problem solving (Hunter & Sparrow, 2012; McCloskey et al., 2009). The term *executive dysfunction* is generally used to describe instances in which these capacities do not function appropriately, which may lead to diminished ability to successfully complete everyday or academic activities and to engage with others in a socially appropriate manner (Anderson, 2008; Slick et al., 2006).

The struggle to identify a concrete definition of EF is related to a number of factors (Hunter & Sparrow, 2012). Specifically, EF is considered a highly complex skill that is influenced by numerous neural networks and structures. In other words, in order for higher-order thinking to occur, frontal regions of the brain must recruit lower-level abilities, such as sensorimotor skills. Therefore, when attempting to measure EF alone, it is difficult to determine whether observed deficits are related to poor EF skills or to underlying deficits in other areas of functioning. Another challenge in understanding EF arises from disagreement between researchers in how best to measure EF. As mentioned previously, a variety of cognitive and behavioral tools exist; yet, these tools often measure different aspects of EF or attempt to measure similar components in different ways, making it difficult to compare performance across tasks (Packwood et al., 2011).

As a result of the aforementioned difficulties, a number of vastly different EF theories have been created. These theories generally fall into one of two categories: unitary or multimodal (Flanagan & Harrison, 2012). Unitary theories view EF as a single construct, similar to an intelligence quotient (IQ), that oversees and guides all lower-level functions (Davis, 2011; Jurado & Rosselli, 2007). In these models, EF is considered an umbrella term that encompasses a diverse system of cognitive processes, which are all driven by one regulatory and goal-directed executive (Barkley, 2012b; McCloskey & Perkins, 2012). Unitary theories have become less popular in recent years but continue to exist due to the high level of inter-correlations between components of EF (Flanagan & Harrison, 2012). Additionally, unitary theories have been supported by studies examining EF skills in young children, which are more limited and less complex than those of adults (Wiebe, Espy, & Charak, 2008). For example, a factor analysis indicated that inhibition, working memory, and planning tend to load on a single factor in children between the ages of four and six (Hughes, Ensor, Wilson, & Graham, 2010).

Multimodal theories of EF break down subcomponents of EF into separate processes, which are thought to function separately (Banich, 2009; Blair & Ursache, 2011). These components often include a large number of skills, which vary significantly between theories. For example, depending upon the theory, EF subcomponents might include initiating, inhibiting, shifting, emotional control, working memory, planning, problem-solving, monitoring, and many others (Aron, 2008; Barkley, 1997; Lezak, 2004). McCloskey and Perkins (2012) offer one of the most comprehensive lists of thirty-three self-regulative EF capacities while Stuss and Benson (1986) define EF as only four components: anticipation, goal selection, pre-planning, and monitoring. One of the most popular definitions of EF is that of Welsh and Pennington (1988), which defines EF as "the ability to maintain an appropriate problem-solving set for attainment of a future goal" (pp. 201-202). This definition includes four elements of EF: intentionality, inhibition, planning, and working memory. The number of EF subcomponents has been reduced and expanded by researchers over time and is continuously changing (Barkley, 2012b; Lezak, 1995; McCloskey & Perkins, 2012).

Baddeley and Hitch's Model of Working Memory

Perhaps the first model to formally outline EF as it is known today is that of Baddeley and Hitch (2000). This model was developed from a well-known model of short-term memory (Baddeley, 1986). Baddeley and Hitch (2000) identify major components of short-term memory, including the phonological loop and visuospatial sketchpad. The phonological loop is responsible for retaining and retrieving verbal information while the visuospatial sketchpad is responsible for retaining and retrieving visual stimuli. These two systems are termed *slave systems* because they are unable to synthesize information on their own and are dependent upon the *central executive*. The central executive coordinates the flow of activity between the slave systems. Recently, the episodic buffer was added to the model as an additional slave system responsible for temporarily storing and integrating information (Neath & Suprenant, 2003).

Although Baddeley and Hitch's model (2000) is primarily related to the processes of storing and encoding information, it recognizes the importance of a system overseeing those processes. It provides a basic framework for working memory, which is considered an aspect of EF by some researchers but not by others (Dehn, 2008). Additionally, Baddley and Hitch's model (2000) implies the importance of selective and sustained attention in memory processes. Although working memory itself is not always considered a component of EF, the model includes a central executive system, which is thought to supervise and regulate attentional control, thereby serving as an EF mechanism. Many researchers have proposed that the central executive as described by Baddeley and Hitch (2000) resides in the same area of the brain as basic EF processes (Poletti, 2009).

Denckla's Theory of Executive Function

Another theorist whose work has helped shape the current understanding of EF is Martha Denckla (1996), who defines EF as a set of processes that control various domains. In her theory, EF serves to prevent behavioral responses, anticipate future actions of self and others, initiate thought and behavior through planning, and direct attention away from distractors that interfere with efficiency. In the academic sector, disruptions in efficiency may result in *producing disabilities*, rather than learning disabilities (Denckla, 1996; McCloskey & Perkins, 2012). Although Denckla (1996) has delineated specific processes that may interfere with domains of functioning, she notes that EF includes every cognitive function regulated by the frontal lobe, which is consistent with the idea that the PFC regulates EF processes (McCloskey & Perkins, 2012; Poletti, 2009). Using research on the nature of frontal lobe functions, Denckla (1996) has added new subcomponents to her theory, including intentionality, which is governed by initiation, sustained attention, and cognitive set shifting. Consistent with Baddeley's (1986, 2000) model, Denckla (1996) emphasizes the influence of memory processes on intentionality.

Lezak's Theory of Executive Function

Lezak (1995, 2004) divides EF into four domains or steps: volition, planning, purposive action, and effective performance. Volition describes the ability for one to make the conscious decision to implement a goal-directed action. In this step, an individual's will guides behavior based on the ability to generate attainable goals (Anderson, 2008). Volition includes awareness of a need or want, identification of related goals, and motivation to begin pursuing those goals. Children with deficits in this area may have difficulty beginning activities despite possessing the ability to complete the activities. Planning, which involves the identification and organization of achievable steps to achieve a goal, is the next step in EF processing (Lezak, 2004). Effective planning includes awareness of barriers to success and the ability to make predictions of what will happen as a goal is pursued. The primary cognitive factors that affect one's ability to engage in planning include sustained attention, working memory, and impulse control (Anderson, 2008). Unlike other theorists, Lezak (2004) recognizes that emotional states, in addition to cognitive states, may influence motivation and planning, particularly in the social sector.

The next step in Lezak's model (2004) is purposive action. This step involves the initiation and maintenance of the steps identified during the planning stage. In order to achieve success during this stage, individuals should utilize inhibition skills and mental flexibility to stop or adjust their proposed course of action. The final step in Lezak's model is effective performance, which describes the ability to evaluate one's performance by monitoring errors and making the appropriate adjustments should mistakes occur. Each EF domain within Lezak's (2004) model includes specific behaviors. EF deficits may be observed during any or all steps of processing. Although Lezak's theory outlines a means of structuring the assessment of EF skills, it neglects key areas of functioning (e.g., working memory) and has not been tested empirically (Anderson, 2008).

Anderson's Executive Control System Model

Similar to the previously mentioned model, the control system model views EF as four distinct processes: attentional control, informational processing, cognitive flexibility, and goal setting (Anderson, 2002). Although the four domains are considered separate entities, Anderson argues that each of the components functions together to create an overall executive system. Anderson's model is considered bidirectional in that each component influences and is influenced by the other components. Additionally, different domains are recruited depending upon the type of task. Each domain is thought to relate to activity in specific regions and neural networks in the frontal lobe; however, the regions and networks often operate together when performing tasks.

The first domain within Anderson's (2002) model is attentional control, which involves the ability to attend to a stimulus, inhibit thoughts and behaviors, and sustain attention to a specific stimulus without distraction. This domain includes the ability to self-monitor, identify, and correct mistakes. Deficits in this area include poor self-control, impulsivity, difficulty completing tasks, and uncorrected errors on tasks. Informational processing includes processing speed and the ability to complete tasks with fluency and efficiency. Poor informational processing often results in low output and delayed response or reaction time. This domain is important in overall EF because it enhances one's ability to carry out other EF skills described in this model. The next area described by Anderson is cognitive flexibility, which involves aspects of working memory and set shifting, including processing different types of information at once, dividing attention, mental and behavioral shifting, and learning from mistakes by creating alternative courses of action. Individuals with deficits in this area might exhibit perseverative behaviors and become stuck when attempting to solve a problem. Lastly, the goal setting domain involves setting goals, planning steps to achieve those goals, and self-starting initiatives or behaviors. Poor goal-setting abilities may result in limited critical thinking capacity, poor problem-solving strategies, and difficulty organizing information effectively.

The Supervisory Activating System Model

The Supervisory Activating System Model (SAS) is a model outlining the role of attention as a regulator of active behaviors (Norman & Shallice, 1986). This model delineates two types of attention that are used for different types of activities: automatic actions and actions requiring deliberate use of resources. Automatic actions occur outside of awareness, such as routine or over-learned behaviors, while deliberate use of resources requires focused attention, planning, decision-making, and monitoring. Deliberate attentional resources are often allocated during unique, difficult, or dangerous situations that do not occur on a regular basis.

Two complementary processes are outlined by Norman and Shallice (1986) to help individuals cope with automatic and deliberate responses. First, contention scheduling is recruited when individuals use automatic responding. This process relies upon schemas to identify programs of action needed to complete common tasks. Although automatic, contention scheduling can be complex in that competing schemas must be ignored or multiple schemas must be recruited simultaneously. In novel situations that do not have well-established schemas, the SAS is activated to exhibit attentional control.

Norman and Shallice's model (1986) was updated by Shallice and Burgess (1996) to emphasize the role of the supervisory system in integrating a wide variety of processes in lower level systems. A three-step procedure is described to better understand how novel situations are handled by the supervisory system. First, one must recruit temporary schemas in order to generate a solution to a problem. Next, an individual must test the selected schema in order to determine if it will be adequate in solving the problem. Last, self-monitoring must take place in order to determine the effectiveness of the schema and whether the process should be adjusted in any way. Overall, the SAS model presents advantages in that it takes multiple EF processes into account, recognizes the importance of attentional control, and aligns with empirical research examining associated neural networks (Shallice & Burgess, 1991). However, this model is considered overly simplistic and difficult to measure through assessment.

Barkley's Self-Regulatory Model

Another major theory of EF is Barkley's self-regulatory model (1997). Barkley defines EF as a meta-construct, comprised of many neuropsychological processes that self-regulate behavior in order to attain goals for the future (Barkley, 2012b). He views EF as hierarchical in nature with behavioral inhibition considered the most fundamental aspect of EF (Barkley, 1997). Behavioral inhibition includes both inhibition of habituated behaviors and the ability to control or ignore interfering information (Anderson, 2008). Inhibition is viewed as central to EF because it provides a delay period during which executive processes can be engaged (Barkley, 1997). Executive processes incorporated during periods of inhibition include working memory, affect regulation, motivation, speech internalization, and reconstitution. Working memory is defined as the ability to retain information about past and present demands in order to achieve a goal. One must use working memory to hold a goal in mind while analyzing, planning for, and completing an activity. Regulation of affect involves the recruitment or control of emotion to either motivate oneself or inhibit behavior. Speech internalization can be described as an inner monologue that enhances the ability to plan, analyze, or self-regulate during problem-solving. This process may also be viewed as a type of verbal working memory. Lastly, reconstitution involves the analysis and synthesis of problems in order to break them down and adjust plans.

Barkley's self-regulation model (1997) was initially developed to better understand children with ADHD. His theory is based on the belief that the primary deficit in the disorder is behavioral inhibition while secondary deficits include various other EF processes (Anderson, 2008). Although initially developed as a framework for ADHD, Barkley's model (1997) can assist in understanding normal development. Specifically, EF skills, including the ability to inhibit thoughts and behavior, develop progressively over time in unison with the prefrontal regions of the brain. Therefore, behavioral inhibition is generally less developed in younger children than in older children or adults.

Hot and Cool Executive Function

Another framework for understanding EF involves dividing the construct into two separate subtypes, one of which is cognitive in nature (i.e., cool or metacognitive EF) and the other of which is emotional in nature (i.e., hot or emotional/motivational EF; Ardila, 2008; Nigg & Casey, 2005; Poletti, 2009). This concept was first introduced by Abelson (1963), who defined cool EF as skills related to intellectual thought and hot EF as skills related to emotion. Specifically, cool EF involves skills measured by traditional cognitive and neuropsychological batteries, such as attention, working memory, shifting, and planning. On the other hand, hot EF involves affective processes that oversee motivation and reward associated behaviors, supported by decision-making and self-regulation (Poletti, 2009). The separation of EF into hot and cool systems does not require that the two systems operate independently. Rather, there is a proposed interplay between the two systems, allowing emotion and motivation to greatly affect attention, decision-making and other EF processes. The importance of the role of delayed gratification in understanding the interplay between hot and cool EF was introduced by Mischel, Shoda, and Rodriguez (1989). This conceptualization aligns with neurocognitive research revealing underlying neural mechanisms that differentially but collaboratively support cool and hot EF. Both types of EF are affected by individual factors, including neural networks, developmental level, and stress (Hunter & Sparrow, 2012).

The theory of hot and cool EF aligns well with the current view of ADHD. For example, the inattentive presentation tends to involve cool EF deficits, whereas the hyperactive/impulsive presentation tends to involve hot EF deficits (Nigg & Casey, 2005; Poletti, 2009). In fact, the theory of hot and cool EF is thought to have arisen from the dual process theory of ADHD, which attributes symptomatology to deficits in both cognitive behavioral inhibition and motivational reward systems (Sonuga-Barke, 2002).

This theory is particularly important to the current study because this study seeks to understand the relationship between cognitive and behavioral measures of EF, which according to the hot and cool EF theory are separate entities.

McCloskey's Theory of Executive Function

McCloskey et al. (2009) identify a large number of processes as EF co-conductors, which work together to engage in higher-order thinking and behavior. This model is considered holarchical, rather than hierarchical, because it recognizes that all aspects of EF are strongly inter-related and serve important roles as individual processes but also as parts of a large system. This theory is particularly complex because the various co-conductors are divided into tiers with lower tiers representing more basic functions, such as somatosensory processing and basic attention and retention. Higher tiers represent more complex functions, such as those related to existential goals, identity, and self-awareness. The tiers included in the model are self-activation, self-regulation, self-control, self-generation, and trans-self-integration. Throughout the lifespan, individuals may move up or down these tiers, rather than progressing through the tiers one-at-a-time.

In the context of EF, the self-regulation tier is considered particularly important because it guides and shapes everyday behavior through executive control. McCloskey et al. (2012) divide the self-regulation tier into separate clusters: attention, engagement, optimization, efficiency, memory, inquiry, and solution. Within these clusters are 33 specific processes used for self-regulation, which act independently and interdependently, and form complex neural networks throughout the brain. Each process may operate differently depending upon the arena in which it is activated, whether that domain be intrapersonal, interpersonal, environmental, or symbolic. In certain circumstances EF abilities may be stronger or more developed than in others. In this way, individuals may exhibit variability in EF skills. Additionally, McCloskey and colleagues (2012) argue that EF processes can be either actively or passively engaged; thus, EF processes can be carried out at a primitive level without the conscious intent to achieve a goal.

The Cattell-Horn-Carroll (CHC) Model

The Cattell-Horn Carroll (CHC) model is a theory of intelligence that has guided the development of multiple cognitive and neuropsychological assessment tools (Flanagan & Harrison, 2012). It integrates the *Gf-Gc* theory of intelligence (Horn & Cattell, 1966) and Carroll's (1997) three-stratum theory of intelligence. *Gf-Gc* theory divides intelligence into two major types: fluid intelligence (i.e., *Gf*) and crystallized intelligence (i.e., *Gc*). Over time, *Gf-Gc* theory has expanded to include multiple factors of intelligence. Carroll's three-stratum theory divides general intelligence into a hierarchical structure made up of narrow and broad abilities. CHC theory combines *Gf-Gc* theory and three-stratum theory to outline how intelligence is comprised of broad abilities, which are made up of narrow abilities. CHC theory is important to the current research study because tools assessing certain broad and narrow abilities, such as working memory and broad attention, are often used to draw conclusions regarding a child's EF skills.

Miller's Integrated SNP/CHC Model

Miller's (2013) Integrated SNP/CHC model provides a framework for organizing and interpreting cognitive and neuropsychological assessment data (Miller, 2013). Cognitive abilities are divided into broad, second-order, and third-order classifications. Executive functions make up one broad classification. Second-order classifications for executive functions include concept recognition and generation, problem-solving, fluid reasoning, and planning, response inhibition, and retrieval fluency. The domain of cognitive facilitators/inhibitors is considered another broad-order classification in Miller's model relevant to the study of EF. Miller defines cognitive facilitators/inhibitors as attentional processes that regulate all other higher-order processes. Attention, working memory, and speed, fluency, and efficiency of cognitive processing are considered second-order classifications of cognitive facilitators/inhibitors in the model. Because Miller recommends specific assessment tools for each classification within his model, this model guided the selection of assessments and subtests to be included the analysis as measures of EF.

History of Executive Function

Although formal study of EF began in the 1870s (Barkley, 2012b), the concept is closely linked to the history of neuropsychology as a whole, which originated with ancient philosophers, such as Plato, who first identified the brain as a potential source of mental processes (Simon, 1972). Plato's beliefs about the importance of the mind emerged as a result of the mind-body problem, which questioned the relationship between thoughts and physical movements or behaviors (Gorton, 1987). Plato argued the mind and body were separate entities that were only united through life, allowing the body to temporarily experience the true reality of the soul or mind. This belief is often termed *dualism* (Schultz & Schultz, 2008). Other theorists supported the notion of *monism,* or the existence of mind and body as a single entity (Moller, 1996).

Understanding the Brain

Rene Descartes, a French philosopher, was key in recognizing the role of brain structures in everyday thought and behavior (Schultz & Schultz, 2008). In the early 1600s, he introduced the theory that the pineal gland plays an essential role in the mind's ability to control the brain. Although the term EF did not arise until much later, Descartes pioneered the idea that a biological control center exists to influence thought and behavior (McCloskey & Perkins, 2012; Poletti, 2009). The work of Thomas Willis, an English doctor, built upon this idea by exploring the anatomy of the brain and nervous system, and coining the term *neurology*. Franz Josef Gall, a German neuroanatomist, then introduced the concept of localized brain functions (Schultz & Schultz, 2008). Specifically, he proposed that the shape of one's skull could determine the quality of his or her mental processes and personality traits. This idea is often termed *cranioscopy* or *phrenology*. Although phrenology has been criticized and is considered pseudoscience, Gall's idea serves as a basis for the attribution of EF processes to the PFC, an example of localized mental processing.

The 1800s brought about discoveries that revolutionized the idea of specialized brain structures, particularly in regards to the frontal regions. Paul Broca and Carl Wernicke were among the first researchers to complete brain lesion studies examining the effects of damage to specific areas of the brain (Kaitaro, 2001). These studies allowed Broca and Wernicke to discover small regions of the left frontal lobe that regulate speech production and comprehension, respectively. While these findings sparked interest within the medical community regarding the specificity of brain structures, it was not until the tragic railroad accident of Phineas Gage (Harlow, 1848) that the importance of specialized brain processing was comprehended at a behavioral level. Gage experienced tremendous damage to his left frontal lobe after a large iron rod passed through his skull. In his case study of Gage, Harlow noted the immense personality and behavioral changes that occurred following Gage's accident, including outbursts of aggression. Hence, Harlow suggested that frontal lobe specialization encompasses more than language through the regulation of behavioral traits, such as impulse-control or inhibition (Aron, 2008), which is now considered a key EF subcomponent (Barkley, 2012b).

Intelligence and Neuropsychological Testing

The history of intelligence testing and neuropsychological assessment also contributed to today's understanding and testing of EF. During the late 1800s and early 1900s, interest grew in attempting to measure various components of thought (Guthrie, 2004). Perhaps most notably, Alfred Binet and Theodore Simon created the first documented intelligence test in an attempt to determine which children should or should not receive education (Binet, Simon, & Town, 1912). Soon thereafter, the idea of determining an intelligence quotient based on testing was introduced by Terman (1922). Initially, neuropsychological testing relied on single tests to determine overall functioning (Miller, 2013). For example, during World War II, the Halstead-Reitan

Neuropsychological Test Battery was created to discriminate between soldiers with and without brain damage (Reitan & Wolfson, 1993). In many cases, soldiers who exhibited extreme head trauma or psychiatric illness were placed in mental institutions or treated with frontal lobotomies in an attempt to reverse changes in personality and behavior (Mashour, Walker, & Martuza, 2005). After white matter tracts in the frontal lobe were severed, patients often displayed flat affect, impaired attention, and disinhibition, which are now considered signs of executive dysfunction (Barkley, 2012b). Thus, it was during this time period that many scientists began to realize the importance of the PFC.

Executive Function

The work of Alexander Luria (1966) is considered foundational to today's understanding of EF. His work addresses the role of the PFC in multiple EF capacities, including self-regulation and planning (Maricle, Johnson, & Avirett, 2010). Luria drew much of his theory from the work of Vladimir Bekhterev, who noted that damage to the frontal lobes resulted in changes in goal-directed behavior (Barkley, 2012b). Based on Bekhterev's findings as well as his own, Luria (1973) concluded that damage to the PFC results in deterioration of the brain's *critical faculty*, a component of the brain that allows an individual to evaluate his or her behavior. Although Luria did not coin the term *executive function*, he identified the frontal lobe as essential in the use and organization of higher-order abilities (Hunter & Sparrow, 2012). When frontal regions are damaged, individuals rely more heavily on automatic behavioral responses than on conscious thought. Luria (1966) used the term *frontal lobe syndrome* to describe deficits associated with frontal lobe impairment. Karl Pribram (1973) was the first theorist to use the term *executive* to describe the functions of the PFC (Barkley, 2012b). With the introduction of this new term, Luria's *frontal lobe syndrome* became known as *executive disorder* and subsequently *dysexecutive syndrome* (Baddeley, 1986; Fuster, 1997). To assist in the identification of this syndrome and other neuropsychological conditions, the Luria-Nebraska Neuropsychological Battery was created in 1978 and revised in 1986 for use in the United States (Golden & Freshwater, 2001). This battery included 14 scales assessing various neuropsychological skills based on quantitative and qualitative data.

Neurobiology of Executive Function

To best understand the neuroanatomical structures and functions related to EF, researchers have studied brain scans of individuals with both average and subaverage EF skills (Davis, 2011). However, it should be noted that abnormalities in specific brain structures cannot solely explain abnormal functioning; instead, self-regulation and other aspects of cognitive functioning are strongly affected by the brain's ability to integrate information flowing between neural networks connecting frontal regions to more primitive areas of the brain (Blair & Ursache, 2011). Additionally, much of the current research exploring human EF is derived from studies involving animals due to ethical constraints, making it difficult to draw concrete conclusions about the neurobiology of EF in humans (Chudasama, 2011). This is a particular concern because animals do not possess many of the same higher-order reasoning capacities as humans.

Brain Regions and Neural Networks

Research investigating EF has consistently supported the notion that the prefrontal cortex (PFC) plays a crucial role in the regulation of cognitive and affective processing through EF capacities (McCloskey et al., 2009; Poletti, 2009). The PFC is located in the most anterior portion of the frontal lobe and can be divided into the dorsolateral PFC, orbitofrontal PFC, ventromedial PFC, and anterior cingulate cortex (ACC). Each of these regions are thought to function together to influence overall EF ability (Flanagan & Harrison, 2012). Together, these components create a complicated network of processes and comprise nearly one-third of the cerebral cortex (Blumenfeld, 2010). Because each region underlies different components of EF, individuals with damage to the PFC may exhibit very different deficits depending upon the particular area in which the damage occurs. Furthermore, these deficits may manifest in certain situations but not in others. For example, one might experience poor self-regulation in social situations but perform well on cognitive measures requiring inhibition or vice versa.

Generally speaking, damage to the PFC has been found to result in poor performance on traditional measures of EF, including the Wisconsin Card Sort and Stroop Color and Word Test, which assess working memory and planning abilities, respectively (Aron, 2008). Additionally, injury to this area has been linked to reductions in goal-directed behavior (Blumenfeld, 2010) and in difficulties inhibiting behavior and engaging in cognitive set shifting (Goldberg, 2001). The role of the PFC is evident even in young children given that studies have revealed an increased level of blood flow to this region in infants engaged in complex searching tasks (Baird, Kagan, Gaudette, Walz, Hershlag, & Boas, 2002).

As EF is considered a highly complex construct, the neuroanatomical and chemical factors underlying the construct are equally if not more complex. Therefore, a variety of regions and networks are thought to contribute to one's ability to self-regulate through EF (Sonuga-Barke, 2002; Zelazo & Carlson, 2012). The neural circuitry within the dorsolateral region of the PFC has been extensively studied in regards to associated deficits in sustained attention, cognitive flexibility, initiation of goal-directed behavior, and other EF factors. (Hale & Fiorello, 2004; McCloskey & Perkins, 2012; Miller, 2013). Rhesus monkeys with damage to the dorsolateral PFC have been found to commonly exhibit poor planning strategies in comparison to peer monkeys without such damage (Chudasama, 2011). Although the dorsolateral PFC is considered the primary region affecting multiple EF abilities, other brain regions, including the parietal, temporal, and occipital lobes, limbic system, and brain stem, mediate these EF abilities through connections to the dorsolateral PFC (Hale & Fiorello, 2004). For example, the hippocampus and cerebellum are recruited during planning tasks (Chudasama, 2011). Damage to various regions of the brain may result in poor EF skills by disrupting networks to and from the dorsolateral PFC (Anderson, 2008).

Differing neural pathways and circuits likely underlie cognitive versus behavioral EF abilities (Zelazo & Carlson, 2012). The dorsolateral PFC is thought to play a role in cognitive aspects of EF (i.e., cool EF), whereas the orbitofrontal cortex and its connections to subcortical regions are thought to underlie emotional aspects of EF (i.e., hot EF). This theory is supported by the case of study of Phineas Gage (Harlow, 1848), who sustained an injury to the OFC, leading to loss of empathy and personality change, including sudden outbursts of anger (Miller, 2013). As was the case with Phineas Gage, the emotional regulatory capacity of the orbitofrontal cortex over the limbic system was reduced, resulting in poor monitoring of emotional responses. Similarly, in children with ADHD, orbitofrontal PFC dysfunction may result in poor behavior inhibition considering that rats with damage to this region often exhibit perseverative responses and behaviors (Logue & Gould, 2014; Chudasama, 2011).

The anterior cingulate cortex is another brain region associated with EF through the regulation of attentional control, response inhibition, working memory, attention, and motivation (Hale & Fiorello, 2004; Miller, 2013; Posner, 1994). Additionally, neural imaging suggests that the anterior cingulate cortex is related to one's abilities to monitor behavior, make goal-directed decisions, and exert cognitive control (Menon & Uddin, 2010). Dysfunction in this region may contribute to apathy, poor attention allocation to academic work, low creativity, perseverative responding, and aversion to delayed rewards, which are characteristic features of ADHD that can lead to negative academic outcomes (Shallice et al., 2002). For example, children diagnosed with ADHD tend to choose large, delayed rewards only 34% of the time, compared to 58% for neurotypical children (Lambek, Tannock, Dalsgaard, Trillingsgaard, Damm, & Thomsen, 2010). The observed difference in delay aversion between groups indicates an underlying ACC deficit in children with ADHD not characteristic of nonclinical populations.

Although regions of the PFC are key facilitators of EF, pathways between the PFC and various subcortical regions of the brain are also involved in self-regulatory processing (Miller, 2013). Cascade theory is a predominant theory used to explain the means through which the brain engages in activities requiring EF (Banich, 2009). This theory relies heavily on neuroimaging studies that have demonstrated a cascade of neural events that occur in a specific order as individuals complete EF tasks. The initial activity occurs in the dorsolateral PFC in order to orient attention to the task at hand and ignore stimuli that is not relevant to the task. Next, the mid-dorsolateral PFC and posterior PFC are engaged to identify key information and select steps for completing a task. The latter area is particularly important in deciding which course of action is best between two competing responses and in recognizing if an error has been made.

The basal ganglia thalamocortical loop further regulates attention, working memory, and inhibition through specific neurotransmitters: dopamine and noradrenaline (Brocki, Fan, & Fossella, 2008). Within this loop, the PFC sends input to the striatum of the basal ganglia, projecting to the planning regions of the putamen and eventually the thalamus, primary motor cortex, and muscles (Aron, 2008). The degree of activation in these networks, as measured by functional magnetic resonance imaging, directly correlates with performance on the stop-signal task, a measure of inhibitory control (Aron, Behrens, Smith, Frank, & Poldrack, 2007). Children with EF dysfunction, including those with ADHD have been found to exhibit decreased grey matter in the structures making up this loop (Qiu, Ye, Li, Liu, Xie, & Wang, 2011; Rappoport, Castellanos, Gogate, Janson, Kohler, & Nelson, 2001). Furthermore, white matter or myelin, an axonal coating that facilitates efficient neurotransmission, lines the tracts between these regions, indicating the importance of rapid communication in EF abilities.

Neurotransmitters

EF skills are strongly related to functioning in the PFC but are mediated by dopamine, serotonin, norepinephrine, and acetylcholine (Logue & Gould, 2014). The

PFC and associated regions contain a large number of pyramidal cells, which play a role in complex higher-order reasoning and are very susceptible to neurochemical changes within the brain (Hosenbocus & Chalal, 2012). Therefore, small disruptions in neurotransmitter regulation (i.e., over or under excitation) may result in abnormalities in a child's ability to self-regulate, interpret rewards, or experience motivation to complete a task. This is particularly true when dopamine levels become disrupted (Logue & Gould, 2014). The medial PFC and orbitofrontal PFC are made up of a large number of specific dopamine receptor sites and receive dopamine projections from the ventral tegmental region of the brain. Dopamine depletion within the medial PFC has been associated with deficits in attention and set-shifting skills whereas excess dopamine has been associated with hyper-attention and enhanced ability to shift between activities or thought processes. However, this effect is not observed when dopamine levels are high or low within the orbitofrontal PFC. In other words, the role of dopamine varies between regions likely because different regions underlie different EF abilities, such as attention, set-shifting, or working memory. Dopaminergic pathways connecting the PFC and basal ganglia are thought to underlie working memory skills by allowing individuals to determine which information to hold in memory and which information to ignore (Banich, 2009).

Norepinephrine is thought to impact EF processes in both the medial PFC and orbitofrontal PFC through projections from the locus coeruleus (Logue & Gould, 2014). Specifically, rat studies have demonstrated that decreases in norepinephrine are associated with impaired attention and set-shifting. Selective norepinephrine reuptake inhibitors, such as amoxetine, have the potential to increase set-shifting abilities by increasing synaptic levels of norepinephrine (Newman, Castonguay, Borkovec, Fisher, & Nordberg, 2008). Additionally, increased levels of norepinephrine in the orbitofrontal cortex is linked to enhanced response inhibition and reversal learning (Logue & Gould, 2014). This neurotransmitter is thought to play a substantial role in modulating EF capacities because it is involved with overall level of arousal and likely determines baseline brain activity when an individual is engaged in tasks requiring EF. In this way, norepinephrine serves as a chemical means of preparing the brain to engage in complex activities.

Serotonin is another key neurotransmitter involved in EF regulation. Projections from the dorsal raphe affect serotonin levels in the PFC (Logue & Gould, 2014). An increase in serotonin levels in the orbitofrontal cortex has been associated with heightened inhibition abilities. For example, individuals with deficient serotonin levels in this region of the brain tend to perform worse on go/no-go tasks than those with higher serotonin levels. Additionally, primate studies have linked serotonin depletion in the orbitofrontal PFC to perseverative behaviors. The role of serotonin in regulating EF is in contrast to that of dopamine. Whereas dopamine affects set-shifting and attention through the medial PFC, serotonin affects response inhibition and reversal learning through the orbitofrontal PFC. This difference exemplifies how different components of EF are regulated by both specific brain regions and neurotransmitters.

Cholinergic input to the medial PFC and orbitofrontal PFC comes from the nucleus basalis of Meynert, which is located in the basal forebrain (Logue & Gould, 2014). The primary neurotransmitter that plays a role in the cholinergic system is acetylcholine. Two types of acetylcholine receptors exist: muscarinic receptors, which are believed to play a role in cognitive flexibility, and nicotinic receptors, which modulate the neurotransmitters involved in EF in the PFC. Nicotine strongly affects one's ability to attend, shift between sets, and inhibit responses. For example, animal studies have demonstrated that increases in nicotine often produce increases of dopamine within the PFC possibly by influencing neuronal firing. It is clear that this system plays a significant role in affecting dopamine levels but research is needed to determine the extent to which cholinergic processes influence norepinephrine and serotonin.

Development of Executive Function

Historically, many theorists argued that EF does not develop until early adulthood; however, recent studies have demonstrated that EF skills begin to develop earlier in life and become more complex over time (Banich, 2009). Specifically, basic EF skills, such as the ability to attend, seem to emerge during the first year of life and develop gradually with key developmental periods occurring between the ages of one and five, seven and nine, and eleven and thirteen (Hughes, 2011; Rhoades et al., 2011). EF subcomponents develop at different times with primitive EF capacities developing during toddlerhood and more complex skills, such as working memory, shifting, and planning consolidating after the age of five (Best, Miller, & Jones, 2009). Accordingly, the cognitive complexity and control theory argues that EF skills become increasingly complex with new rules governing goal-directed behavior across development (Zelazo, Muller, Frye, & Marcovitch, 2003).

Theories of Executive Function Development

Specific theories of EF development include the extended phenotype theory of EF (Barkley, 2012b) and the holarchical model of EF capacities (McCloskey et al., 2009). In these models, EF skills build upon one another to encompass new cognitive abilities and more complex behaviors over time. For example, the first level of the extended phenotype theory (Barkley, 2012b), the pre-executive level, includes basic sensory processes, such as language, memory, and motor functioning, which are over-ridden by more complex processes in order to carry out goal-directed actions. Likewise, the first three levels of the holarchical level involve the mastery of basic self-control through activation, regulation, realization, and determination (McCloskey et al., 2009). The highest levels of both frameworks utilize primitive processes, such as self-motivation and attention, to integrate the mind and body into a sense of self (McCloskey & Perkins, 2012) and to adhere to abstract principles of ethics, morality, and law within the social context (Barkley, 2012b). The developmental progression through increasingly complex levels of EF mirrors the maturation of the PFC from the posterior to the anterior regions (Hughes, 2011).

Executive Function Throughout the Lifespan

Although specific theories of EF development exist, the concept is more easily discussed from a general perspective. During the first year of life, infants develop the ability to attend to their world in order to have their needs met (Hunter & Sparrow, 2012). Over time, they begin to explore their worlds with increased awareness while engaging basic cognitive flexibility and problem-solving to experiment with the effects their behaviors have on their environment. These early experiences are key to development because they allow children to form schemas that are used during later stages of development in goal-setting and problem-solving. According to Garon, Bryson, and Smith (2008), the toddler period involves the most rapid changes in EF skills. Anywhere from one to four new EF abilities, including working memory, set shifting, and inhibition, are thought to emerge during this timeframe, particularly between the ages of three and five (Miller, Nevado-Montenegro, & Hinshaw, 2012; Willoughby, Wirth, & Blair, 2012). In later stages of development, EF skills are thought to build upon the skills that initially developed during the infancy and toddler years (Hunter & Sparrow, 2012). Specifically, cognitive flexibility, planning, and basic decision-making arises during middle childhood while foundational EF skills continue to grow. It is during this stage that differences in EF abilities between children become apparent in academic and social settings likely due to increased demands to maintain attention over long periods of time.

Adolescence is considered another critical period of EF development (Hunter & Sparrow, 2012). New EF skills emerge and previously acquired skills are refined. These changes co-occur with synaptic proliferation and pruning in the PFC, which are strongly influenced by hormonal changes associated with puberty. During the adolescent period, individuals become better able to engage in tasks requiring attention, working memory, and problem-solving at a fluent rate and to think critically about their own decisions and behaviors. Despite growth in cognitive EF, development of behavioral EF, which is regulated by the orbitofrontal cortex, tends to develop more slowly (Poletti, 2009; Zelazo & Carlson, 2012). While performance on the Stroop Color and Word Test, which measures cognitive EF skills, improves from childhood to the age of seventeen, performance on the Iowa Gambling Task, which taps risk- and reward-related decision-making, remains poor into adulthood (Zelazo, Qu, & Kesek, 2010). Because adolescents tend to show little prefrontal control over the circuitry involved in emotion and motivation, they tend to seek more immediate and extreme rewards (Blackmore & Choudhury, 2006) and to engage in risky behaviors despite understanding the consequences of such behaviors (Crone, 2009; Zelazo & Carlson, 2012). Moreover, during adolescence, dopaminergic activity is at its greatest (Sisk & Zehr, 2005), which may also account for reward-seeking behavior during this stage.

Peaks in behavioral inhibition, working memory, problem-solving, and set-shifting occur during early adulthood (Hunter & Sparrow, 2012). During these years of peak functioning, executive skills are often easily engaged without conscious effort on behalf of the individual. However, as individuals become older, EF skills tend to become less automatic and again must be engaged deliberately. Declines in working memory and processing speed are strongly associated with increased age. Alzheimer's disease and other types of dementia are often characterized by significant deficits in EF (Banich, 2009). Thus, EF skills follow a specific pattern of growth, developing slowly over time, peaking, and then declining slowly with age.

Neuroanatomical Factors

Overall, the development of EF is strongly related to physical development and, as with physical development, the rate and pattern of growth differs between individuals (Hunter & Sparrow, 2012). Researchers have observed surges in EF that often coincide with physical growth spurts and neural proliferation in the PFC (Jurado & Rosselli, 2007). For example, the PFC undergoes significant changes during puberty and adolescence, which is often when more advanced EF skills reach peak development (i.e., attain adult levels of maturity; Huttenlocher, 1979). Overall, the PFC develops slower than other regions of the brain and myelination in this area continues to occur into young adulthood; thus, EF skills continue to change throughout a significant portion of one's life (Blair & Ursache, 2011).

Brain plasticity is a key concept related to development patterns in EF (Zelazo & Carlson, 2012). Plasticity is defined as sensitive periods during which the brain is more susceptible to change through environmental experiences and stimulation. The brain is considered more malleable during the preschool years and adolescent years, which are two key periods of EF development. During these stages, it is vital for children to experience enriching experiences that boost their EF capacity and to avoid chronic stress and environments that do not allow them to practice their developing skills.

From a neuroanatomical standpoint, synaptic changes are vitally important in EF development. Synapses, or connections between neurons, proliferate in the PFC during early childhood and again during adolescence, eventually leading to synaptic pruning, or elimination of unimportant connections, which improves transmission efficiency (Rakic, 2009). Pruning continues to occur during adolescence, reducing synaptic density in the frontal lobe (Blackmore & Choudhury, 2006). In this way, the development of synapses follows a U-shaped curve of development, with peak synaptic density observed between ages eight and twelve. White matter volume, on the other hand, linearly increases during childhood and adolescence, with older children having greater amounts of myelination in the frontal and parietal regions than younger children. Concurrently, performance on tests of inhibition, working memory, and problem-solving improves throughout adolescence, revealing neurobiological roots of certain EF skills in white matter tracts (Anderson, Anderson, Northam, Jacobs, & Catroppa, 2001; Poletti, 2009). Prefrontal activation patterns also vary between developmental stages with younger individuals demonstrating more widespread activation than adults during EF-related tasks (Casey et al., 1997). Over-activation of the frontal regions may result in poor response inhibition because synaptic pruning has not yet taken place, leading to recruitment of widespread neural networks, rather than efficient, specialized focal regions (Poletti, 2009).

Biological Factors

EF development is strongly influenced by genetic factors. Estimates of the degree to which genetics contribute to EF skills range from 40 to 80% (Leve et al., 2013). Genetic factors may be up to 99% responsible for certain aspects of EF, including inhibition and shifting (Friedman et al., 2008). A twin study of inhibition, working memory, and shifting revealed a strong correlation between these skills after controlling for intelligence; consequently, certain EF abilities may be influenced by genetic factors beyond inherited intelligence.

The genetic factors believed to contribute most significantly to EF are those regulating neurotransmitter activity in the frontal lobes, such as catechol-O-methyltransferase (COMT) and dopamine (Logue & Gould, 2014). COMT is a type of enzyme that affects synaptic levels of dopamine. Multiple COMT alleles, which are alternative forms of a gene that may arise from mutation, are known to exist, including the Val and Met alleles. Individuals with the Val allele may exhibit lower dopamine levels as a result of increased COMT, while individuals with the Met allele may exhibit higher dopamine levels as a result of decreased COMT. As previously mentioned, higher dopamine levels are often associated with stronger EF skills; therefore, individuals with two copies of the Met allele tend to demonstrate stronger attentional control and cognitive flexibility than those with two copies of the Val allele. Gender effects related to these alleles are also observed. Heterozygosity (i.e., one copy of each type of allele) does not appear to enhance working memory performance in females when compared to other females with homozygosity (i.e., two copies of the Val allele). On the other hand, males with one copy of each allele exhibit stronger performance when compared to males with two copies of the Val allele.

Additionally, disruptions in the reelin gene, which place individuals at risk of developing various disorders, including depression, autism spectrum disorder, and schizophrenia, are thought to underlie certain EF deficits (Baune et al., 2010). Certain mutations, or single nucleotide polymorphisms (SNPs), in the reelin gene often result in abnormalities in reeling, a protein associated with neural development. The medial PFC, where certain EF processes are housed, is particularly susceptible to subtle changes in the reelin protein. Depending upon the specific type of SNP, disruptions in reelin protein may result in either increased or decreased EF capacity; thus, this protein is considered key in understanding executive dysfunction as well as the ability to effectively engage in EF-related tasks.

Recent research investigating EF as well as other complex cognitive processes consistently supports the idea that environmental factors influence the expression of genes through the concept of epigenetics (Champagne, 2010). The processes of methylation and acetylation are capable of affecting deoxyribonucleic acid (DNA) in way that activates or inactivates genes. In this way, genetic predisposition to executive dysfunction does not guarantee that one will demonstrate poor EF skills. Thus, it is important to consider the environmental factors that may result in the expression of underlying gene mutations predisposing one to EF disruptions.

Environmental Factors

Multiple environmental factors contribute to the development of EF deficits, including prenatal/birth history and parenting style. Regarding prenatal history, maternal depression and exposure to teratogens have been associated with EF deficits (Hughes et al., 2012). Specifically, the level of maternal depression corresponds with the extent of executive dysfunction, which is most apparent in the areas of inhibition and working memory. For example, mild depression during pregnancy may be associated with less severe deficits than those occurring in connection with severe depression during pregnancy. Additionally, drug and alcohol use during pregnancy have been linked to disruptions in EF (Riccio et al., 2010). Fetal alcohol syndrome (FAS) is characterized by numerous EF deficits, including poor attention, working memory, and reasoning as well as impulsivity. Prenatal exposures to environment toxins, such as lead and mercury, appear to affect EF by altering brain structures and chemistry (Dawson & Guare, 2010). Exposure to pesticides affects the cholinergic system of the brain, resulting in abnormalities in attention and inhibition processes. Furthermore, babies exposed to the aforementioned substances and chemicals during pregnancy are more likely to be born preterm with a low birth weight, which places them at further risk of developing EF abnormalities across multiple domains (Riccio et al., 2010; Ritter, Nelle, Perrig, Steinlin, & Everts, 2013).

Harsh rearing, neglectful parenting, and low cognitive stimulation have also been associated with poor EF skills that persist into adulthood (Rhoades et al., 2011). Disorganized and stressful home environments may inhibit the emergence of EF skills through the child's modeling of parental activity (Hughes, 2011). Furthermore, genetic vulnerability to EF deficits may be expressed only in children exposed to environmental chaos. Positive and responsive interactions with parents reduce the potential for epigenetic changes that lead to the emergence of EF deficits (Rhoades et al., 2011). Such interactions also help children realize that they can influence their environment, increasing their motivation to exert control over their lives through the implementation of EF. Environmental influences may reduce abnormal neurodevelopment associated with ADHD (Halperin & Healey, 2011). Specifically, cognitive stimulation arising from directed play and physical exercise promote positive brain growth, thereby possessing the potential to alter the trajectory of ADHD symptomatology.

The role of parenting styles in EF development can be discussed within the framework of Vygotsky's (1962) sociocultural perspective. Carlson (2003) argues that EF development is highly influenced by certain dimensions of parent-child interactions, two of which coincide with the theories of Vygotsky (1962). Firstly, Carlson (2003) advocates for the use of scaffolding, which is a method introduced by Vygotsky for facilitating the development of skills through the provision of varying degrees of adult guidance. Indeed, Hughes and Ensor (2009) found that maternal scaffolding positively predicts scores on EF measures in preschool age children. Specifically, scaffolding may be used to encourage the development of goal-directed cognition (Hughes, 2011). The second dimension contributing to EF development is mindfulness, which fosters self-regulation through the use of external and internal self-talk (Carlson, 2003). Vygotsky (1962) emphasized the importance of language in the development of tools necessary to interact effectively with one’s world. From this perspective, self-regulation through EF capacities can best be established through the use of inner speech. From an evolutionary perspective, Ardila (2008) also argues that metacognitive EF is deeply rooted in language and therefore plays a role in oral comprehension, reading, and writing.

Cultural Factors

In addition to environmental factors, cultural factors, including socioeconomic status (SES), race, and gender, are known to influence EF. While multiple cultural factors have been found to contribute to a unitary EF construct, the most consistent predictor of EF is SES (Noble et al., 2007). SES, which significantly influences multiple cognitive and neurobiological factors, can be defined as a measure of sociocultural factors that delineate one’s overall status in society (Hackman, Farah, & Meaney, 2010). Although SES affects neurodevelopment at all levels, the most extreme effects are observed within the lowest strata (Adler & Rehkopf, 2008). As SES decreases, concurrent reductions in intelligence and school achievement often occur (Bradley & Corwyn, 2002).

In regards to specific neurocognitive systems, SES contributes most significantly to language processing and two subcomponents of EF: working memory and cognitive control (Noble et al., 2007). Researchers have noted SES-related differences in working memory and inhibition in children as young as six months of age that continue to exist throughout childhood (Lipina, Martelli, Vuelta, & Colombo, 2005; Noble et al., 2007). EF deficits attributed to SES are observed in conjunction with disrupted brain activation patterns, such as low PFC recruitment when confronted with distracting stimuli, which suggests poor attentional control (Kishiyama, Boyce, Jimenez, Perry, & Knight, 2009). Hackman et al. (2010) argue that EF deficits related to SES arise epigenetically with genomic variation in serotonin and dopamine sensitivity predisposing individuals to such deficits, which are then expressed when triggered by conditions of poverty. However, the social causation hypothesis posits that these effects are at least partly reversible through early intervention (Blair & Diamond, 2008).

Environmental variables that may interact with SES and EF include prenatal factors, parenting styles, and cognitive stimulation (Hackman et al., 2010). Low SES can lead to higher rates of stress, infection, and poor nutrition in pregnant women, which can serve as environmental triggers for epigenetic changes through the release of stress hormones, such as corticotropin-releasing factor and glucocorticoids (Meaney, Szyf, & Seckl, 2007; Seckl & Holmes, 2007). Release of these hormones during pregnancy has been associated with inhibition of neurogenesis, or the production of new neurons, and myelination, or the proliferation of white matter tracts, with resulting symptoms of inattention in offspring (Seckl, 2008). Poverty may also increase maternal irritability, depression, and anxiety, which can lead to neurodevelopmental changes in offspring, such as decreased synaptic plasticity, or the rewiring of neural connections, in the prefrontal cortex, interfering with executive control over learning and memory (Hackman et al., 2010). Finally, low SES homes often provide less stimulating environments than high SES homes, undermining the child's ability to develop tools vital for reading and mathematics achievement. However, intervention programs targeting cognitive stimulation may diminish the effects of poverty on cognitive development by promoting school readiness (Reynolds, 1994).

The intersection between race and SES is also an important consideration given recent findings that growing up in a White household may serve as a buffer against the negative effects of poverty on EF development in children (Rhoades et al., 2011). Specifically, growing up in a high-risk household (i.e., a low SES, single-parent household) predisposes children to EF deficits, but the effect of household is more extreme for African American children than for White children. In African American children, poverty leads to deleterious effects on EF and other cognitive capacities unless numerous protective factors are in place, including dual parenting, maternal engagement, and low levels of stress. Unfortunately, the cultural context in which White privilege and classism continue to exist may only exacerbate stress and thus further impair EF in marginalized cultural groups (McIntosh, 1998).

Clinical Diagnoses and Executive Function

Various childhood disorders involve EF deficits, including but not limited to ADHD, autism spectrum disorder, TBI, learning disabilities, depression, and anxiety

(Gioia et al., 2002). Each of these disorders presents with a unique profile of EF deficits (Penningtion & Ozonoff, 1996). ADHD is a neurodevelopmental disorder characterized by difficulty paying attention, excessive activity, and poor behavioral control (American Psychiatric Association [APA], 2013). Certain theorists describe ADHD as an EF disorder due to the central role executive dysfunction plays in the hyperactive and inattentive behaviors observed in children with the disorder (Barkley, 1997). However, researchers' opinions about the central underlying deficit in the disorder vary. Barkley (1997) argues that inhibition is responsible for all deficits observed in children with ADHD, including working memory, affect regulation, motivation, and problem-solving. Children are unable to regulate these processes due to poor cognitive and behavioral inhibition. Similarly, Bayliss and Roodenrys (2000) attribute ADHD symptomatology to a supervisory attentional system lacking inhibitory control. Pennington and Ozonoff (1996) also recognize the importance of inhibition in ADHD but identify working memory as another key area of disruption.

Children with ADHD commonly struggle with academic achievement, which may be attributed to executive dysfunction. For example, children with ADHD and concurrent EF deficits show higher levels of inattention and school problems and score lower on intelligence scales than children with ADHD in the absence of EF deficits (Bierman, Greenberg, Blair, & Domitrovich, 2008). Specifically, deficits in working memory and inhibitory control seem to contribute to the classroom behavior problems frequently demonstrated by children with ADHD (Molfese, Molfese, Molfese, Rudasill, Armstrong, & Starkey, 2010). Behavioral EF may further contribute to school problems by impairing motivation to succeed (Poletti, 2009). Self-motivation is essential in the classroom given that internal demand to produce schoolwork is significantly more effective than external demand from teachers or parents (McCloskey & Perkins, 2012). In addition to EF, processing speed and spatial ability have been found to predict school performance in mathematics when controlling for overall intelligence, which accounts for between 51% and 75% of variance in academic performance (Rohde & Thompson, 2007). Therefore, EF is only one of many neurocognitive factors that contribute to school performance. Even still, findings that children with ADHD tend to demonstrate poor achievement in reading and math, high use of special education services, and low high school graduation rates (Rutledge, van den Bos, McClure, & Schweitzer, 2012) may be partially explained by EF deficits, revealing the need for targeted EF interventions designed to reduce negative outcomes.

Autism spectrum disorder is a neurodevelopmental disorder involving impairments in social communication as well as restricted interests and repetitive behaviors (APA, 2013). There is controversy regarding the role of EF as a causal factor in the disorder; however, most theorists agree that EF deficits tend to co-occur with the disorder (Liss et al., 2001; Russell, 1997). Specifically, individuals with autism tend to exhibit rigid thought patterns related to poor flexibility (Ozonoff & Cathcart, 1998). Additionally, planning and organizational deficits are commonly observed in individuals with autism, including those with high-functioning autism, which often involves verbal disorganization. Theorists have attributed these deficits to "weak central coherence" and poor informational processing (Frith & Happe, 1994; Minshew, Goldstein, & Siegel, 1997). Deficits in metacognition and initiation may partially explain poor adaptive functioning often observed in this population (Gilotty et al., 2002). Working memory skills tend to vary in individuals with autism. While verbal working memory skills may be weak, spatial working memory skills may be average or above average. EF deficits become more pronounced as individuals with autism mature; thus, older children with autism may stand out more as a result of deficits in comparison to same-age peers than younger children with autism (Rosenthal, Wallace, Lawson, Dixon, Yerys, & Kenworthy, 2013).

Traumatic brain injury (TBI) occurs when an external force causes damage to the brain and can vary in severity; thus, EF deficits associated with TBI can vary significantly based on the location and extent of damage (Gioia et al., 2002). Injuries that occur during early periods of development may result in more significant EF disruptions by preventing mastery of EF skills (Riccio et al., 2010) and executive dysfunction often persists many years after the injury (Dawson & Guare, 2010). Deficits in attention, memory, processing speed, flexibility, inhibition, planning, organization, and self-regulation have been observed in individuals with TBI (Anderson, Catroppa, Morse, Haritou, & Rosenfeld, 2000; Dennis, Guger, Roncandin, Barnes, & Schachar, 2001). Generally, disruptions in EF as a result of TBI can be attributed to changes in frontal lobe systems, including damage to white matter tracts. Damage to the dorsolateral PFC, orbitofrontal PFC, and anterior temporal lobes may also place individuals at great risk of executive dysfunction (Levin et al., 1993). As a result of deficits attributed to TBI, individuals with TBI tend to perform and behave best in structured environments (Ganesalingam, Yeates, Taylor, Walz, Stancin, & Wade, 2011).

Learning disabilities, particularly reading disorders, may relate to underlying deficits in certain EF abilities (Gioia et al., 2002). Specifically, working memory has been found to contribute to deficits in word recognition and reading comprehension (De Jong, 1998; Swanson, 1999; Swanson & Ashbaker, 2000). Additional deficits in planning and organizing tend to present in individuals with reading comprehension difficulties (Levin, 1990). Children with learning disabilities in the area of math tend to exhibit EF deficits in the areas of problem-solving, logical reasoning, and working memory (Riccio et al., 2010). It is likely that the struggle to filter out distracting information and select and switch between strategies contributes to problems in mathematic skills. Additionally, the classroom environment presents novel activities and assignments each day, requiring ample amounts of attention devoted to problem-solving, which may explain the role of EF in other learning disabilities (Diaz & Berk, 1992). Motivation, which is considered a component of EF, may also be disrupted, resulting in difficulty initiating school-related activities and projects (Barkley, 2012b). In other words, in the classroom, children with EF deficits may exhibit symptoms preventing them from achieving academic success, including poor motivation to succeed, self-starting behavior, attention to homework assignments, and problem-solving skills.

Individuals with depressive and anxiety disorders may also exhibit some level executive dysfunction (Hosenbocus & Chahal, 2012). Specifically, depression has been linked to decreased levels of dopamine, which is also associated with impairments in EF.

This neurochemical abnormality may provide an explanation as to why individuals with depression often exhibit poor emotional and behavioral regulation. Furthermore, due to decreased levels of energy, those with depression tend to have difficulty initiating or sustaining attention to tasks, which may be conceptualized as a deficit in EF. Some researchers even consider suicidal ideation an EF deficit given that individuals who consider suicide often lack the reasoning abilities to identify alternative solutions to their problems (Keilp, Wyatt, Gorlyn, Oquendo, Burke, & Mann, 2014). Less research examining the link between EF and anxiety disorders is available. However, there is some evidence that obsessive thoughts and compulsive behaviors associated with obsessive-compulsive disorder relate to underlying deficits in cognitive shifting and inhibition (Hunter & Sparrow, 2012). Furthermore, stress levels and sleep difficulties as a result of anxiety may contribute to impaired EF (Meaney et al., 2007).

Assessment of Executive Function

Researchers and clinicians have found it challenging to assess EF with validity and reliability due to a number of reasons, including problems with the definition of EF, performance-based testing, test environment, and ecological validity (Gioia et al., 2002). First, there is vast disagreement regarding how best to define the construct of EF. Definitions that do exist are considered theoretical, rather than operational (Hughes & Graham, 2002). Thus, a variety of instruments exist to assess different aspects of EF in different ways. EF is a complex and multifaceted construct by nature. Therefore, individuals may perform well on certain measures of EF but poorly on other measures. This has been evidenced in studies examining performance in individuals with frontal lobe damage (Eslinger & Damasio, 1985; Shallice & Burgess, 1991). Furthermore, because both higher- and lower-order processes are recruited when engaged in EF-related tasks, aspects of functioning may be overlooked when attempting to measure the construct (Lezak, 1995). In other words, an individual may display poor performance on a task thought to measure one area of functioning when the deficit is actually in a lower-order ability, such as visual or motor processing. This problem is often referred to as task impurity (Banich, 2009). In order to avoid completing inaccurate and limited assessments, it is important for clinicians to recognize the integration of skills that is necessary when an individual completes a task requiring EF (Gioia & Isquith, 2004).

Another barrier to effective assessment of EF involves test environment. Many neuropsychological assessments are based on performance in an unnatural setting for a short period of time (Pennington & Ozonoff, 1996; Shallice & Burgess, 1991). Because test settings are often very quiet and free from distractions, EF deficits may not be demonstrated as they would be in a more typical setting (Cripe, 1996; Lezak, 1982). Additionally, certain EF skills, such as setting goals, planning, and multi-tasking are less necessary because test instructions often control these factors for examinees, thereby relieving EF demands (Manchester, Priestley, & Jackson, 2004). These concerns elicit an additional concern regarding ecological validity of test results. Because of the conditions in which neuropsychological testing takes place, results are unlikely to generalize to everyday situations and environments (Silver, 2000). In other words, ecological validity, or the degree to which performance coincides with real-world functioning, is likely low when measuring EF (Slick et al., 2006). Accordingly, Burgess, Alderman, Evans, Emslie, and Wilson (1998) found very low correlations between children's scores on EF tasks and behavioral ratings by their caregivers.

Cognitive Measures of Executive Function

Cognitive or neuropsychological measures of EF include classic tests that were not originally developed to assess EF but are now considered valid measures of the concept. Classic tools thought to measure EF include the Wisconsin Cart Sorting Test (WCST; Heaton et al., 1993), Stroop Color and Word Test (Golden & Freshwater, 2002), as well as trail-making tests, go/no-go tests, and tower tests (Davis, 2011). The WCST requires examinees to sort 64 cards by color, number, and shape while adhering to a set of sorting rules (Heaton et al., 1993). This assessment is viewed as a gold standard of EF assessment tools and measures set-shifting, working memory, problem-solving, and attention. The Stroop Interference Test was developed in 1935 and was later renamed Stroop Color and Word Test (Golden & Freshwater, 2002). To complete this task, individuals must use selective attention and cognitive flexibility to read color words printed in different colors and to then name the color of the ink the word is printed in on a separate trial. Trail making tests generally assess shifting skills, requiring examinees to draw lines from one stimulus to another in multiple trials with different stimuli and rules (Reitan & Wolfson, 1993). Go-no/go tasks vary across measures but generally assess set-shifting, attention, and inhibition (Davis, 2011). Lastly, tower tests, such as the Tower of Hanoi and Tower of London, assess problem-solving and cognitive flexibility. Examines must move disks on a pegboard in the lowest possible amount of moves while following specific rules in order to match the disks and pegboard to a target picture.

Many of the aforementioned tools were developed as a means of assessing overall brain functioning following brain damage (Heaton et al., 1993) and have recently been adapted and featured as subtests in neuropsychological tests, including the NEPSY-2, D-KEFS, and TEA-Ch. These tests incorporate the Boston Process Approach, which was developed by Kaplan (1988) in an attempt to understand the mental processes individuals recruit to identify answers to complex problems or questions. In testing situations, qualitative behaviors are observed and recorded in order to assist the clinician in understanding thought patterns (Miller, 2013). By observing qualitative behaviors that are not normally taken into account when completing a cognitive assessment, clinicians are better able to draw conclusions about a child's functioning in his or her natural setting.

The D-KEFS (Delis et al., 2001) is a battery of tests intended to measure EF in both children and adults. Rather than being developed alongside a theory, the D-KEFS was originally designed to evaluate verbal and nonverbal higher-order or frontal lobe abilities in children and adults who had experienced mild brain damage. Many of the tests included in the D-KEFS were derived from numerous traditional neuropsychological tests, including the Tower of Hanoi (Delis et al., 2001) and Stroop Color and Word Test (Golden & Freshwater, 2002). It is comprised of nine individual tests, including tests adapted from the Tower of Hanoi and Stroop Color and Word Test. Multiple scores are available to assess various aspects of EF, including set shifting, inhibition, concept generation, planning, and reasoning. The D-KEFS was not designed with the intention of investigating EF holistically; instead, it was developed as means of including various EF tasks in one instrument. Therefore, tasks may measure the same or different aspects of EF and some may be more reliable and valid than others. A more thorough discussion of the D-KEFS reliability and validity will be included in chapter 3.

The NEPSY-2 (Korkman et al., 2007) is a neuropsychological battery used to assess various neurocognitive skills in children. The battery includes 32 subtests measuring six domains of functioning: memory and learning, visuospatial abilities, language, sensorimotor functioning, social perception, and attention and executive function. Examiners may create individualized batteries by selecting the most appropriate subtests to administer to each test subject. Test publishers encourage examiners to administer a variety of subtests in order to rule out lower-order deficits that may underlie disruptions in higher-order functioning. Tests included in the attention and executive function domain of the NEPSY-2 measure sustained and selective attention, concept generation, set shifting, and working memory. A more thorough discussion of the NEPSY-2 reliability and validity will be provided in chapter 3.

The TEA-Ch (Manly et al.,1999) is comprised of nine subtests measuring selective attention, sustained attention, and attentional shifting in children. Subtests are presented in a "game-like" format and are designed to reduce demands of memory, motor speed, and language in order to more accurately assess attention. The test has been used to effectively discriminate between children with and without ADHD (Heaton et al., 1993). In addition to assessing aspects of attention, certain subtests within the TEA-Ch provide information regarding set shifting and inhibitory control. The TEA-Ch has recently been revised and published as the Test of Everyday Attention for Children-Second Edition (TEA-Ch-2; Manly, Anderson, Crawford, George, Underbjerg, &

Robertson, 2016); however, it is not yet available for use in the United States. For the purposes of this study, the TEA-Ch remains relevant due to the recency of the revised version's publication and corresponding lack of necessary data and research. A more thorough discussion of the TEA-Ch reliability and validity can be found in Chapter 3.

Intelligence measures may also be used to assess EF abilities. These tests include the Kaufman Assessment Battery for Children, Second Edition (KABC-2) and Wechsler Intelligence Scale for Children, Fifth Edition (Wechsler, 2014). Of particular relevance is the Woodcock Johnson Tests of Cognitive Abilities, Third Edition (WJ III COG NU), which utilizes CHC theory to conceptualize intellectual abilities (McGrew et al., 2007). The WJ III COG NU provides subtests that assess aspects of attentional control, reasoning skills, working memory, and planning. These subtests are included in the Executive Processes clinical cluster, Broad Attention cluster, and Fluid Reasoning (i.e., Gf) cluster. The Woodcock Johnson Tests of Cognitive Abilities, Fourth Edition (WJ IV COG; Schrank, McGrew, & Mather, 2014) was recently published; however, the test is so new that there is inadequate data available in the dataset to justify its use in the current study. Additionally, many of the subtests included in the WJ IV COG are very similar to those included in the WJ III COG. A more thorough discussion of WJ III COG reliability and validity can be found in Chapter 3.

Behavioral Ratings of Executive Function

Behavioral rating scales have also been created as a means of attempting to assess EF abilities in an individual's everyday environment (Gioia et al., 2002). A recent meta-analysis indicates that behavioral measures of EF identify different underlying skills than those identified by cognitive or neuropsychological measures of EF (Toplak, West, & Stanovich, 2013). In comparison to cognitive measures, behavioral rating scales tend to have high levels of ecological validity because they more accurately predict behaviors in natural settings (Franzen & Wilhelm, 1996; McCloskey et al., 2009). The strongest and most ecologically valid behavioral ratings are drawn from multiple raters, including parents, other caregivers, and teachers (Gioia et al., 2000). However, behavioral scales can be significantly limited when rater bias exists, resulting in under- or over-reporting of behaviors. Raters only see children in certain environments, making it difficult to generalize results across settings unless multiple raters from home and school environments complete forms.

Several EF ratings scales have been made available in recent years, including the BRIEF (Gioia et al., 2000) Barkley Deficits in Executive Functioning Scale: Children and Adolescents (BDEFS: CA, Barkley, 2012a), Delis-Rating of Executive Functions (D-REF; Delis, 2012), and Comprehensive Executive Function Inventory (CEFI, Naglieri & Goldstein, 2012). Each of these scales uses a questionnaire format to produce different scores indicating functioning within different domains of EF. The BRIEF is the only behavioral rating scale that will be used as part of the current study; therefore, the other behavioral rating scales will not be reviewed here. The BRIEF is available for preschool, school-age, and adolescent children in caregiver, teacher, and self-report forms depending upon the age of the child. Eight clinical scales are available and measure different aspects of EF: Inhibit, Shift, Emotional Control, Initiate, Working Memory, Plan/Organize, Organization of Materials, and Monitor. These scales load onto two empirically validated indices: Behavioral Regulation and Metacognition. Additional validity indices are also provided to determine whether responses are overly negative or inconsistent. The BRIEF-2 (Gioia, Isquith, Guy, & Kenworthy, 2015) was recently published; however, the authors of the BRIEF-2 did not change any of the items that contributed to the composite scores. Because the structure of the test did not change, the original BRIEF is easily compared to the BRIEF-2. A review of BRIEF reliability and validity will be provided in chapter 3.

The Behavior Assessment System for Children, Second Edition (BASC-2; Reynolds & Kamphaus, 2004) is another behavioral rating scale that can be used to assist in the identification of executive dysfunction. Although a newer version of this system is available, the second edition will be used as part of the current study and will therefore be reviewed here. Like the BRIEF, the BASC-2 is available for different ages and raters and provides scores on various scales that load on broader factors. The BASC-2 does include an Executive Functioning scale; however, scores on this content scale were not available for use in the current study. Instead, content scales assessing behaviors that may indicate underlying EF disruptions will be used. These scales such include hyperactivity, aggression, conduct problems, learning problems, attention problems, atypicality, and others. A review of BASC-2 reliability and validity will be provided in chapter 3.

Summary

This literature review highlights the crucial role of EF as it relates to everyday functioning for individuals with and without clinical diagnoses. Research on this topic has accelerated in recent years, leading to a large number of theories regarding EF; yet a single widely-accepted model has not been agreed upon. Most theorists agree that EF is composed of various subcomponents that function both independently and interdependently. EF skills are primarily affected by neurochemical processes within the prefrontal cortex of the brain but lower-level systems are recruited to support performance on different types of EF-related tasks. Multiple biological, environmental, and cultural factors influence one's ability to effectively engage in activities requiring EF. Because of the lack of a clear and consistent definition as well as the complex nature of EF, it is difficult to assess EF with validity and reliability. The current study aimed to contribute to the field by investigating the relationship between cognitive and behavioral measures of EF using a variety of assessment tools believed to target EF skills based on CHC and Integrated SNP/CHC theories. Research design and methodology will be discussed in the following chapter.

CHAPTER III

METHODS

This chapter outlines the methods utilized in the current study, which was designed to explore the relationship between performance on cognitive measures of executive function (EF) and parent ratings of EF in a mixed clinical sample. The primary purpose of this study was to determine whether parent ratings of EF, derived from parental impressions of child behaviors, were similar to scores obtained on instruments designed to measure performance-based or cognitive EF. This chapter will also provide an overview of the participants, procedures, measures, and statistical analyses utilized.

Research Participants

Data was culled from an archival database of neuropsychological case studies submitted by students of the School Neuropsychology Post-Graduate Certification Program (KIDS, Inc). The archival database consisted of over 1600 case studies collected between 2006 and 2016. A number of exclusion criteria were used to select cases from the larger dataset. First, case studies collected within the last five years (2011-2016; *n* = approximately 600) were reviewed first, primarily because these cases were the most comprehensive and contained more complete data relevant to the current study. Cases were also removed if a diagnostic category was not indicated, if the participant's age fell outside of the range of 8 to 16 years, and if there were no data provided for both the BRIEF and BASC-2. The final sample consisted of 176 participants. These case studies included information on children with a variety of diagnoses, including ADHD, autism, emotional disturbance, specific learning disabilities, and general neurological impairments or medical conditions (e.g., seizures, TBI, and brain tumors). Individuals ranging in age from 8 to 16 years and with a variety of racial and ethnic backgrounds were included in the study. More detailed descriptive statistics are provided in Chapter 4.

Procedures

As previously mentioned, archived data collected between 2011 and 2016 were utilized for data analysis. Permission to access the clinical dataset was obtained from Dr. Daniel Miller, Executive Director of KIDS, Inc. Case studies contained in the dataset were selected based on the availability of scores on behavioral and cognitive measures of EF. As mentioned previously, cases were initially sorted and selected based on exclusion criteria, including the presence of the required instruments/subtests proposed for the study (BRIEF, BASC-2, D-KEFS, TEA-Ch, WJ III COG, and NEPSY-2). After the dataset was narrowed down to 176 participants, the variables were examined to determine whether adequate numbers of cases contained data for each variable. Variables with missing data for over one-third of participants were excluded from the analysis with the exception of the Activities of Daily Living scale on the BASC-2. A large number of variables originally anticipated to be included in the analysis were removed in order to reduce the amount of missing data within the sample. These variables included all subtests of the TEA-Ch, Sorting Condition 1, Sorting Condition 2, and Sorting Confirmed Correct on the D-KEFS, and Auditory Attention, Auditory Working Memory, and Planning on the WJ III COG NU.

Due to the clinical nature of the dataset, it was anticipated that no case would have 100% of the requisite variables. In other words, inconsistency between referral batteries resulted in significant variation in assessments administered and, thus, high rates of data points missing at random. Multiple imputation (MI) was initially identified as a means of addressing missing data points. However, very little research has been conducted examining the use of factor analysis with datasets that have undergone MI; therefore, full information maximum likelihood (FIML) and pairwise deletion were selected as more appropriate alternatives. FIML utilizes available data to estimate parameter values that maximize the likelihood of obtaining the data within the dataset while pairwise deletion minimizes loss by allowing cases containing some data to be included in the analysis (Collins, Shafer, & Kam, 2001). FIML was utilized during the factor analysis stage of the study and pairwise deletion was utilized during the correlation and regression phases of the study.

Measures

The Integrated School Neuropsychology/Cattell-Horn-Carroll Model (Integrated SNP/CHC model) was used to identify behavioral and cognitive measures of EF (Miller, 2013). The Integrated SNP/CHC model is a framework for understanding neurocognitive functioning that recognizes various domains of processing, including EF. Certain subtests from commonly used assessment tools have been identified as targeting specific skills within each neurocognitive domain. Subtests representing the EF domain were selected for use in this study depending upon the availability of data within the dataset and, thus, may not correspond exactly with the Integrated SNP/CHC model. The following paragraphs provide an overview of the assessment tools and subtests utilized, including descriptions of each subtest as well as reliability and validity data for each tool.

Behavior Assessment System for Children, Second Edition (BASC-2)

The BASC-2 (Reynolds & Kamphaus, 2004) is a multi-method rating system that evaluates student behavior and self-perceptions in children and young adults between the ages of 2 and 25. The system is considered multi-method because rating scales are available in teacher, parent, and self-report forms along with a developmental history questionnaire and an observation recording system. The BASC-2 Parent Rating Scales (PRS) were utilized for the purposes of the current study in order to assess parent perceptions of child behaviors that may indicate underlying EF dysfunction. One rater for each child was selected for use, including mother, father, and other caregivers. For cases in which data was available for more than one rater, information provided by the mother was included. The PRS has three forms available: preschool, child, and adolescent. Both the child and adolescent forms were included in the current study. Although the child and adolescent forms utilize some different questions, they are believed to measure the same constructs and are therefore considered comparable. The PRS utilizes a four-choice format in which the rater selects one of four choices to describe the frequency of a behavior, ranging from *Never* to *Almost Always*. These ratings assess various areas of clinical and adaptive functioning.

The BASC-2 PRS provides a variety of composite, primary, and content scores. The BASC-2 utilizes *T* scores to indicate the severity of various behaviors. On the clinical scales, higher *T* scores indicate more maladaptive functioning. On the adaptive scales, lower *T* scores correspond to more maladaptive functioning. *T* scores have an average of 50 and a standard deviation of 10; therefore, scores ranging from 30 to 40 on the adaptive scales or 60 to 70 on the clinical scales are considered at-risk and scores falling below 30 on the adaptive scales or above 70 on the clinical scales are considered clinically significant. The following clinical scales were selected for use in the current study: Activities of Daily Living, Adaptability, Aggression, Anxiety, Attention Problems, Atypicality, Conduct Problems, Depression, Functional Communication, Hyperactivity, Leadership, Social Skills, Somatization, and Withdrawal. See Table 3.1 for subtest descriptions. A BASC-2 EF scale exists but these scores were not included in the dataset and so is unavailable for use in the current study.

Normative data. The BASC-2 provides comparison of scores to various norm groups, including General, Clinical, Learning Disability, and Attention-Deficit/Hyperactivity Disorder (Reynolds & Kamphaus, 2004). The normative data was drawn from a large national sample of children in public and private classrooms representing the population of children in the United States in the areas of race/ethnicity, parent education level, special education classification, and geographic region. The norms are divided by age and therefore, each score is compared to students falling within the same age range. Examiners can also select whether to compare students' scores to combined or separate sex norms; however, to maximize the number of cases included in the current study, cases using both combined and separate norms were utilized. The publishers note that when combined-sex norms are used, males and females may show slightly higher scores in certain areas (e.g., males higher in Aggression or females higher in Social Skills).

Reliability and validity. Studies have been completed to demonstrate the internal consistency, test-retest, and interrater reliability of the BASC-2 PRS (Reynolds & Kamphaus, 2004). In the General Norm group, internal consistency scores for individual scale scores were found to be high, ranging from .83 to .87 for the child and adolescent scales. Of the scales included in the current study, Hyperactivity and Attention Problems were found to have the highest internal consistency ratings. In order to examine test-retest reliability, parents rated their child twice with an interval of 9 to 70 days between completion of the rating scales. Median test-retest reliability values for the individual scales were high at .84 and .81 for the child and adolescent scales, respectively. Finally, two parents or caregivers were asked to rate their child at similar times (i.e., between 0 and 70 days of each other). Median inter-rater reliabilities were found to be .69 to .77 for the child and adolescent scales, respectively.

The validity of the BASC-2 PRS is demonstrated through scale intercorrelations and factor analysis, comparison of scores to other behavioral measures, profiles of children with specific classifications, and response validity indicators (Reynolds & Kamphaus, 2004). Regarding intercorrelations, scales expected to be related showed high correlations, including Hyperactivity, Aggression, and Conduct problems, whereas those expected to be dissimilar did not demonstrate strong correlations. Additionally, individual scales making up the same composite scores were found to be strongly correlated.

Table 3.1
BASC-2 Scores Included in the Analysis

Test – Composite Score	Description
BASC-2 – Activities of Daily Living	Skills associated with performing basic tasks in an acceptable and safe manner.
BASC-2 – Adaptability	Ability to adapt readily to changes in the environment.
BASC-2 – Aggression	Tendency to act in a hostile manner (either verbal or physical) that is threatening to others.
BASC-2 – Anxiety	Tendency to be nervous, fearful, or worried about real or imagined problems.
BASC-2 – Attention Problems	Tendency to be easily distracted and unable to concentrate more than momentarily.
BASC-2 – Atypicality	Tendency to behave in ways that are considered “odd” or commonly associated with psychosis.
BASC-2 – Conduct Problems	Tendency to engage in antisocial and rule-breaking behavior.
BASC-2 – Depression	Feelings of unhappiness, sadness, and stress that may result in an inability to carry out everyday activities or may bring on thoughts of suicide.
BASC-2 – Functional Communication	Ability to express ideas and communicate in a way others can easily understand.
BASC-2 – Hyperactivity	Tendency to be overly active, rush through work or activities, and act without thinking.
BASC-2 – Leadership	Skills associated with accomplishing academic, social, or community goals, including the ability to work with others.
BASC-2 – Social Skills	Skills necessary for interacting successfully with peers and adults in home, school, and community settings.
BASC-2 – Somatization	Tendency to be overly sensitive to and complain about relatively minor physical problems and discomforts.
BASC-2 – Withdrawal	Tendency to evade others to avoid social contact.

Note. Adapted from *BASC-2: Behavior assessment system for children-second edition* by C. Reynolds and R. Kamphaus. Copyright 2004 by NCS Pearson, Inc.

Confirmatory factor analysis and principal-axis factoring support the existence of a factor structure in which the individual scales are divided into four or five composite scores, including Externalizing and Internalizing. These composite scores are available but were not used as part of the current study to avoid the simultaneous use of individual subtest or scaled scores and composite scores. The use of both can result in statistical error because the individual scores contribute to the composite scores.

The BASC-2 was also compared to a variety of behavioral scales: Achenbach System of Empirically Based Assessment Child Behavior Checklist (ASEBA; Achenbach & Rescorla, 2001), Conners' Parent Rating Scale-Revised (CPRS-R; Conners, 1997), BRIEF (Gioia et al., 2000), and the original BASC Parent Rating Scales. Similarly named individual scores on the ASEBA and BASC-2 tended to correlate at a moderate to high level (Reynolds & Kamphaus, 2004). The same was found when comparing the BASC-2 to the CPRS-R and BRIEF. Hyperactivity, Atypicality, and Attention Problems on the BASC-2 had the strongest correlations with scores included in the Behavioral Regulation Index on the BRIEF, supporting their inclusion in the current study. Validity indicators are also available on the BASC-2 PRS. First, the Response Pattern index identifies patterns that may indicate the respondent was inattentive to the content of the items (e.g., identical responding to multiple items in a row or in a cyclical pattern). Additionally, the Consistency index indicates whether the respondent has provided differing responses to items that are commonly rated similarly.

Scale descriptions. The following scales were selected for use in the current study because children with EF dysfunction may display some or many of the behaviors included in each of the scales (Barkley, 2012a; McCloskey & Perkins, 2012; Weyandt, 2005; Molfese et al., 2010). The Activities of Daily Living scale measures one's ability to perform everyday tasks safely and appropriately (Reynolds & Kamphaus, 2004). Internal consistency ranged from .70 to .76 for the child and adolescent forms while test-retest reliability was .86 (child) and .82 (adolescent) and interrater reliability was .80 (child) and .86 (adolescent). The SEM value for this scale was 5.1 for both child and adolescent.

The Adaptability scale measures one's ability to adapt to changes in his or her environment (Reynolds & Kamphaus, 2004). Internal consistency ranged from .81 to .86 for the child and adolescent forms while test-retest reliability was .84 (child) and .74 (adolescent) and interrater reliability was .82 (child) and .78 (adolescent). The SEM values for this scale were 4.3 (child) and 4.1 (adolescent).

The Aggression scale measures one's tendency to behave in a verbally or physically threatening manner to others (Reynolds & Kamphaus, 2004). Items measuring this scale include those assessing arguing, making threats, calling names, breaking other's possessions, or hitting others. Internal consistency ranged from .84 to .89 for the child and adolescent forms while test-retest reliability was .72 (child) and .84 (adolescent) and interrater reliability was .58 (child) and .79 (adolescent). The SEM value for this scale was 3.8 for both child and adolescent.

The Anxiety scale measures the tendency to experience nervousness, fearfulness, or worry about real or imagined problems (Reynolds & Kamphaus, 2004). Internal consistency ranged from .80 to .86 for the child and adolescent forms while test-retest reliability was .65 (child) and .86 (adolescent) and interrater reliability was .86 (child) and .69 (adolescent). The SEM values for this scale were 4.0 (child) and 4.3 (adolescent).

The Attention Problems scale provides information regarding one's tendency to have difficulty concentrating or to become distracted (Reynolds & Kamphaus, 2004). Items assess common characteristics of ADHD, which is sometimes considered a disorder of EF (Barkley, 2012b). Internal consistency ranged from .83 to .90 for the child and adolescent forms while test-retest reliability was .87 (child) and .84 (adolescent) and interrater reliability was .80 for both child and adolescent forms (Reynolds & Kamphaus, 2004). The SEM values for this scale were 3.8 (child) and 3.6 (adolescent).

The Atypicality scale assesses unusual behaviors that are viewed by others as odd or that may indicate psychosis (Reynolds & Kamphaus, 2004). These behaviors include being disconnected or unaware of one's surroundings, acting confused, or saying things that do not make sense. Internal consistency ranged from .77 to .86 for the child and adolescent forms while test-retest reliability was .80 (child) and .82 (adolescent) and interrater reliability was .69 (child) and .71 (adolescent). The SEM values for this scale were 4.0 (child) and 4.6 (adolescent).

The Conduct Problems scale measures the tendency for one to engage in rule-breaking or antisocial behaviors, including cheating, lying, using drugs or alcohol, running away, or stealing (Reynolds & Kamphaus, 2004). Internal consistency ranged from .82 to .89 for the child and adolescent forms while test-retest reliability was .85 for child and adolescent and interrater reliability was .65 (child) and .79 (adolescent). The SEM values for this scale were 3.8 (child) and 3.6 (adolescent).

The Depression scale provides information related to mood, including unhappiness, difficulty carrying out everyday activities, and suicidality (Reynolds & Kamphaus, 2004). Internal consistency ranged from .85 to .88 for the child and adolescent forms while test-retest reliability was .85 (child) and .87 (adolescent) and interrater reliability was .77 (child) and .86 (adolescent). The SEM values for this scale were 3.6 (child) and 3.7 (adolescent).

The Functional Communication scale measures one's ability to express his or her ideas in an understandable manner (Reynolds & Kamphaus, 2004). Internal consistency ranged from .83 to .88 for the child and adolescent forms while test-retest reliability was .84 (child) and .82 (adolescent) and interrater reliability was .82 for both child and adolescent versions. The SEM values for this scale were 3.9 (child) and 3.6 (adolescent).

The Hyperactivity scale measures over-activity and impulsivity, including the tendency to rush through schoolwork, interrupt others, or behave without thinking (Reynolds & Kamphaus, 2004). Internal consistency ranged from .76 to .87 for the child and adolescent forms while test-retest reliability was .89 (child) and .74 (adolescent) and interrater reliability was .74 (child) and .78 (adolescent). The SEM values for this scale were 3.8 (child) and 4.4 (adolescent).

The Leadership scale provides information regarding one's ability to work with others to accomplish academic, social, and community goals (Reynolds & Kamphaus, 2004). Internal consistency ranged from .81 to .86 for the child and adolescent forms while test-retest reliability was .86 (child) and .81 (adolescent) and interrater reliability was .72 (child) and .83 (adolescent). The SEM values for this scale were 4.1 (child) and 3.8 (adolescent).

The Social Skills scale provides information regarding the ability to engage with others in a socially appropriate manner across various settings (Reynolds & Kamphaus, 2004). Internal consistency ranged from .83 to .88 for the child and adolescent forms while test-retest reliability was .84 (child) and .77 (adolescent) and interrater reliability was .75 (child) and .72 (adolescent). The SEM values for this scale were 3.8 (child) and 3.6 (adolescent).

The Somatization scale measures the tendency to be sensitive to minor physical problems and discomfort (Reynolds & Kamphaus, 2004). Internal consistency ranged from .77 to .83 for the child and adolescent forms while test-retest reliability was .66 (child) and .86 (adolescent) and interrater reliability was .58 (child) and .67 (adolescent).

Finally, the Withdrawal scale measures the extent to which one tends to avoid social contact with others (Reynolds & Kamphaus, 2004). Internal consistency ranged from .76 to .83 for the child and adolescent forms while test-retest reliability was .83 (child) and .78 (adolescent) and interrater reliability was .70 (child) and .81 (adolescent). The SEM values for this scale were 4.6 (child) and 4.4 (adolescent).

Behavior Rating Inventory of Executive Function (BRIEF)

The BRIEF is a rating scale designed to assess behaviors related to EF in school age children between the ages of 5 and 18 (Gioia et al., 2000). The BRIEF was developed based on the theory that EF is composed of subdomains of regulatory functions, including initiating behavior, inhibiting actions, selecting goals, planning and organizing as a means of solving problems, shifting between strategies, and monitoring performance or behavior (Donders, 2002). Thus, the inventory provides scores on eight clinical scales, which are drawn from either parent or teacher responses to 86 items, in order to assess a variety of behaviors that may indicate EF dysfunction (Gioia et al., 2000). The clinical scales are divided into the following subtypes of EF: Inhibit, Shift, Emotional Control, Initiate, Working Memory, Plan/Organize, Organization of Materials, and Monitor. Responses across the scales also come together to form two broad indices (i.e., Behavioral Regulation and Metacognition) as well as a Global Executive Composite, which provides an overall EF score. The eight subscale scores obtained from the Parent Form were included in the analysis for the current study (see Table 3.2). As with the BASC-2, the BRIEF utilizes *T* scores to determine the child's level of EF dysfunction with higher scores indicating greater levels of difficulty (Gioia et al., 2000). Scores falling above 65 are considered clinically significant.

Normative data. The general norm group for the BRIEF Parent Form includes 1,419 cases. A clinical sample is also available and includes children with various neurodevelopmental disorders, such as ADHD (Gioia et al., 2000). Normative data was drawn from students in public and private schools in the state of Maryland. The norm group was sampled with the goal of approximating the United States population according to gender, socioeconomic status, race/ethnicity, and age. Separate norms by gender were indicated based on significant differences across most scaled scores by gender. Additionally, because unique developmental trends were found by age, norms were developed for separate age groupings (5 to 7, 8 to 10, 11 to 13, and 14 to 18 years).

Overall, scores on the BRIEF are positively skewed with scores clustering towards the lower end.

Table 3.2

BRIEF Scores Included in the Analysis

Test – Composite Score	Description
BRIEF – Inhibit	Assesses inhibitory control (i.e., the ability to inhibit, resist, or not act on an impulse) and the ability to stop one's own behavior at the appropriate time.
BRIEF – Shift	Assesses the ability to move freely from one situation, activity, or aspect of a problem to another as the circumstances demand.
BRIEF – Emotional Control	Addresses the manifestation of executive functions within the emotional realm and assesses a child's ability to modulate emotional responses.
BRIEF – Initiate	Contains items relating to beginning a task or activity, as well as independently generating ideas, responses, or problem-solving strategies.
BRIEF – Working Memory	Measures the capacity to hold information in mind for the purpose of completing a task.
BRIEF – Plan/Organize	Measures the child's ability to manage current and future-oriented task demands.
BRIEF – Organization of Materials	Measures orderliness of work, play, and storage spaces (e.g., such as desks, lockers, backpacks, and bedrooms).
BRIEF – Monitor	Assesses work-checking habits (i.e., whether a child assesses his or her own performance during or shortly after finishing a task to ensure appropriate attainment of a goal).

Note. Adapted from *BRIEF: Behavior rating inventory of executive function* by G. Gioia, P. Isquith, S. Guy, and L. Kenworthy. Copyright 2000 by Psychological Assessment Resources, Inc.

Reliability and validity. Evidence of reliability for the BRIEF is demonstrated through internal consistency, interrater reliability, and test-retest reliability values (Gioia et al., 2000). The BRIEF has high internal consistency, ranging from .80 to .98 for Parent and Teacher forms. Interrater reliability between parents and teachers was found to be moderate with a mean value of .32. However, low correlations were obtained on Initiate (r = .18) and Organization of Materials (r = .15). The test authors note that these differences may be attributed to differences in structure between the home and school environments. Overall, parents reported significantly more problems than did teachers across scales. Test-retest correlations for the parent normative sample ranged from .76 to .85 with an average interval of 2 weeks between completion of scales.

To establish construct validity for the BRIEF, twelve pediatric neuropsychologists were asked to examine certain items on the BRIEF and determine whether they represented the intended EF domains or constructs (Gioia et al., 2000). Items with poor interrater agreement were not included in the final version of the test. To demonstrate construct validity, the BRIEF was then compared to a variety of existing scales, the ADHD-Rating Scale-IV (ADHD-IV; DuPaul, Power, Anastopoulos, & Reid, 1998), Child Behavior Checklist (CBCL; Achenbach, 1991a), Teacher's Report Form (TRF; Achenbach, 1991b), Behavior Assessment System for Children (BASC; Reynolds & Kamphaus, 1992), and Conners' Rating Scale (CRS; Conners, 1989). The results of these studies were consistent with the idea that specific EF deficits should affect behaviors in certain domains (Gioia et al., 2000). For example, the ADHD-IV Hyperactivity-Impulsivity scale correlated with the BRIEF Inhibit (r = .73) scale while the Behavioral

Regulation Index scores on the BRIEF correlated strongly with the Aggression ($r = .76$) and Hyperactivity ($r = .63$) scales on the BASC. Additionally, the TRF Aggressive Behavior scale correlated with the BRIEF Inhibit ($r = .83$), Shift ($r = .70$), Emotional Control ($r = .81$), and Monitor ($r = .74$) scales. Scales targeting attention problems tended to correlate with the Initiate, Working Memory, and Plan/Organize scales on the BRIEF across measures. More recently, the concurrent validity of this instrument has been established through the comparison of the clinical scales of the BRIEF to performance-based measures of EF, including classic tests of set-shifting and inhibition (Toplak, Bucciarelli, Jain, & Tannock, 2009).

To further establish the validity of the BRIEF, factor analyses were completed (Gioia et al., 2000). Principal axis factoring supported the existence of a two-factor model, which was divided into Metacognition and Behavioral Regulation. Principal axis factoring was also completed with the CBCL, BASC, and ADHD-IV scales included. These studies demonstrated that the BRIEF scales Working Memory and Inhibit generally converged with other scales of inattention and impulsivity, respectively. Additionally, behavioral and emotional factors, such as aggression and emotional control were shown to diverge from other factors of EF. These results suggest that a variety of similar but differing EF factors exist.

Finally, studies examining the profiles of children belonging to various clinical groups have shown that differing EF profiles emerge for certain groups, including TBI and learning disorders, suggesting the BRIEF possesses some discriminatory ability (Gioia et al., 2000). Most significantly, the BRIEF appears to have strong predictive validity in regards to ADHD. Parent ratings on the Working Memory and Inhibit scales exhibit sensitivity to detecting children with a likely diagnosis of ADHD, particularly with inattentive symptoms.

Scale descriptions. Each of the individual scales on the BRIEF were included in the analysis for the current study. The individual scales included in the broad Metacognition Index are Initiate, Working Memory, Plan/Organize, Organization of Materials, and Monitor (Gioia et al., 2000). The Initiate scale provides information on the ability to independently begin an activity or generate an idea or strategy. Children with deficits in this area may require excessive prompts to begin an activity or have difficulty fluently retrieving words from memory. Internal consistency was .80 while test-retest reliability was .80 and interrater reliability was .18. The Working Memory scale assesses the ability to hold information in awareness and carry out multistep activities using that information. Children with poor working memory skills may have difficulty remembering information or may forget what they are doing while in the process of completing a task. There is strong overlap between this scale and measures of attention. Internal consistency was .89 while test-retest reliability was .85 and interrater reliability was .30. The Plan/Organize scale evaluates the ability to anticipate events, set goals, and determine the steps necessary to obtain goals as well as the ability to communicate information accurately by organizing thoughts. Deficits in this area may manifest as difficulty preparing for an assignment in a timely manner or becoming overwhelmed by large amounts of information. Internal consistency was .90 while test-retest reliability was .85 and interrater reliability was .35. The Organization of Materials scale measures the physical orderliness of a child's work and living spaces. Children with poor physical organization often struggle in school as a result of not having necessary materials readily available. Internal consistency was .87 while test-retest reliability was .79 and interrater reliability was .15. Finally, the Monitor scale is aimed at assessing one's ability to check his or her performance to reduce careless mistakes. It also includes the ability to recognize the effect of one's behavior upon others. Internal consistency was .83 while test-retest reliability was .76 and interrater reliability was .42.

The individual scales included in the broad Behavioral Regulation Index of the BRIEF are Inhibit, Shift, and Emotional Control (Gioia et al., 2000). The Inhibit scale is designed to assess the ability to resist acting impulsively or automatically. Behaviors may include intrusiveness, lack of personal safety, tendency to interrupt others, and high levels of physical activity. Internal consistency was .91 for the Parent Form normative sample while test-retest reliability was .84 and interrater reliability was .50. The Shift scale assesses the ability to easily move between one situation, activity, or strategy to another. Deficits may result in difficulty switching attention, solving problems flexibly, and making transitions. Children may demonstrate perseverative behaviors and require consistent routines. Internal consistency was .81 while test-retest reliability was .78 and interrater reliability was .15. The Emotional Control scale provides a measure of one's ability to modulate emotional responses. Emotional lability or explosiveness may indicate poor emotional control. Children exhibiting a deficit in this area may have excessive emotional reactions to minor upsets. Internal consistency was .89 while test-retest reliability was .79 and interrater reliability was .18.

Delis Kaplan Executive Function System (D-KEFS)

The D-KEFS (Delis et al., 2001) is a comprehensive battery of tests designed to assess EF in children and adults between the ages of 8 and 89 (with the exception of the subtest Proverbs, which can only be administered to individuals age 16 and older). The battery is comprised of nine individual tests that may be administered individually or together. When administered as an entire battery, the D-KEFS is approximately 90 minutes long. The D-KEFS does not provide an overall executive function score or composite scores. Instead, examiners are allowed to select the tests to administer and may administer them in any order, interpreting the scores on an individual basis.

Scores provided encompass a variety of performance measures, including total achievement on a task, total correct responses, condition scores, and error scores. Most of the scores provided by the D-KEFS are scaled scores, which have a mean of 10 and a standard deviation of 3; however, certain subtests utilize cumulate percentile ranks. Many subtests include multiple trials or conditions, which become increasingly complex and, therefore, require greater cognitive complexity as the examinee progresses through the trials. Additionally, the D-KEFS provides optional analysis of qualitative data, including omission and commission errors, which may indicate perseveration or other difficulties. Initially, the following scaled scores from the following ten subtests were expected to be included in the analysis: Color Word Interference Conditions 3 and 4, Verbal Fluency Condition 3, Design Fluency Condition 3, Trail Making Condition 4, Sorting Condition 1, 2, and Confirmed Correct Sorts, Tower Total Achievement, and Word Context Total Consecutively Correct. Each of the Sorting scores were excluded due to insufficient amounts of available data. Table 3.3 lists the subtests from the D-KEFS that were initially selected for inclusion.

Normative data. The D-KEFS normative sample consists of 1,750 individuals between the ages of 8 and 89 years of age (Delis et al., 2001). Of the participants included, 875 ranged in age from 8 to 19 years. The sample was stratified by sex, race, ethnicity, educational level, and geographic region to reflect the demographics of the United States population according to 2000 Census data. The majority of subtest scaled scores were normally distributed; however, certain measures, particularly those analyzing error rates, were skewed in the negative direction. Negatively skewed scaled scores were converted to cumulative percentages. Studies of normative data also indicated age effects across most tasks. Children between the ages of 8 and 10 demonstrated steep rises in raw scores, particularly on tasks measuring verbal skills (e.g., Verbal Fluency), while adults between the ages of 70 and 89 demonstrated significant decline in raw scores. These age effects are reflected in the standard deviations of each subtest. In other words, the standard deviations are much larger in the youngest and oldest age groups than in age groups in between. Delis et al. (2001) suggest these differences indicate variability in rates of cognitive development and aging, particularly in the area of executive function.

Reliability and validity. Evidence of reliability exists in the availability of internal consistency, test-retest reliability, and standard error of measurement (SEM; Delis et al., 2001). Split-half procedures were utilized to provide reliability coefficients of internal consistency. Because the D-KEFS provides subtests that measure both basic and higher-order skills in repeated trials, internal consistency ratings are negatively impacted.

Studies of test-retest reliability included a sample of 101 cases encompassing all age groups with an average of 25 days between administrations. In general, performance tended to improve slightly over time. Finally, SEM was used to determine the degree of certainty that a subject's true scores fell within a certain range or confidence interval using internal consistency and test-retest coefficients.

Table 3.3

D-KEFS Scores Included in the Analysis

Test – Composite Score	Area of EF Measured
D-KEFS – Color Word Interference 3	Set Shifting
D-KEFS – Color Word Interference 4	Set Shifting, Inhibition
D-KEFS – Verbal Fluency Condition 3	Set Shifting
D-KEFS – Design Fluency Condition 3	Set Shifting
D-KEFS – Trail Making Condition 4	Set Shifting
D-KEFS – Tower Total Achievement	Planning, Inhibition
D-KEFS – Word Context Total	Deductive Reasoning

Note. Adapted from *Delis-Kaplan Executive Function System Examiner's Manual* by D. C. Delis, E. Kaplan, & J. H. Kramer. Copyright 2001 by The Psychological Corporation.

The majority of the subtests on the D-KEFS are new versions or modifications of long-standing tests (Delis et al., 2001). The validity of the original tests have been well-established through numerous studies investigating their sensitivity in detecting brain damage, particularly in the frontal region (Lezak, 1995; Spreen & Strauss, 1991). Certain subtests on the D-KEFS (i.e., Color-Word Interference, Trail Making, Tower, Word Context, Design Fluency, and Verbal Fluency) have been shown to discriminate between certain clinical groups, including those with fetal alcohol syndrome and frontal lesions (Mattson, Goodman, Caine, Delis, & Riley, 1999; Baldo, Shimamura, Delis, Kramer, & Kaplan, 2001). Construct validity of the D-KEFS is supported by observed correlations between subtests indicating the tasks measure similar constructs. In other words, tasks evaluating similar components of EF tend to correlate. Finally, divergent validity has been evidenced by low to moderate correlations between tasks on the D-KEFS and other assessments including the California Verbal Learning Test-Second Edition (CVLT-II; Delis, Kramer, Kaplan, & Ober, 2000) and the Wisconsin Card Sorting Test (Berg, 1948; Heaton et al., 1993).

Subtest descriptions. The Color Word Interference task assumes similarities to the traditional Stroop Color and Word Test (Golden & Freshwater, 2002). It incorporates multiple conditions in which the examinee is asked to quickly provide the name of a series of colors (Condition 1), then to read the printed names for each of the colors (Condition 2), then to name the ink color of the printed names (Condition 3), and finally to accurately switch between the tasks of Condition 2 and Condition 3 (Condition 4; Delis et al., 2001). The Condition 3 and 4 scores were utilized in the current study as a means of assessing verbal inhibition and attentional switching based on Miller's theoretical model (2013). These scores are derived from total completion time. Internal consistency and SEM values were not provided for Condition 3 and 4 scores but test-retest reliability was found to be high at .90.

Verbal Fluency is also considered a measure of verbal attentional shifting (Miller, 2013). This subtest requires the student to complete three conditions: quickly retrieving words that sound phonetically similar (Condition 1), quickly retrieving words that are semantically similar (Condition 2), and quickly switching between retrieving words falling in the aforementioned categories (Condition 3; Delis et al., 2001). The Condition 3 score, which measures total number correct, was included in the data analysis for the current study as a measure of EF. For children and adolescents, internal consistency scores on this condition ranged from .37 to .62, indicating a small to medium amount of reliability. Similarly, the test-retest reliability coefficient was .65. Finally, the SEM values ranged from 1.48 to 2.06 for subjects between the ages of 8 and 19.

The Design Fluency task also utilizes multiple conditions and assesses visual attentional switching abilities (Miller, 2013). In order to complete this task, the examinee must create novel connections between a series of dots according to a set of rules. First, the examinee is presented with an array of solid dots and is asked to quickly create as many novel designs as possible in a certain time frame by connecting the solid dots (Condition 1). Next, the examinee is presented with both solid and empty dots and is asked to create as many novel designs as possible by connecting only empty dots (Condition 2). Lastly, Condition 3 requires that the child alternate between connecting solid and empty dots while creating novel designs (Condition 3). The Condition 3 score was included in the data analysis for the current study. Internal consistency values were not made available because item interdependence prevented the use of these procedures. Test-retest reliability was low on Condition 3 with a correlation of only 0.13. The SEM value on this condition for all ages was 2.47.

The Trail Making subtest is a modification of the original Trail Making Test, which was developed in 1938 (Partington & Leiter, 1949). It includes 5 conditions measuring various aspects of cognition. The examinee must fluently utilize visual scanning and EF to identify target stimuli (Condition 1), connect numbers in numerical order (Condition 2), connect letters in alphabetical order (Condition 3), alternate between connecting numbers and letters in order (Condition 4), and accurately trace a line across a page (Condition 5). Scores are derived from completion time. Conditions 1, 2, 3, and 5 allow the examiner to determine if a low score on Condition 4 indicates actual deficits in cognitive shifting or if performance should be attributed to delays in motor or sequencing abilities. The Condition 4 score was utilized in the current study as a measure of visual attentional shifting, an aspect of EF processing (Miller, 2013). Internal consistency values for Condition 4 were not provided and test-retest reliability was found to be low (.20).

Sorting is another subtest modified from a traditional neuropsychological measure, measuring concept formation and generation (Miller, 2013). Examinees are asked to sort cards picturing different shapes, words, and patterns into two separate piles with an equal number of cards in each pile. They must then explain their reasoning for sorting the cards in that manner. This task is completed multiple times with two different sets of cards, providing overall scores for each set (Conditions 1 and 2). These scores are determined based on whether the sorts are correct as well as the level of appropriateness and accuracy with which the examinee describes the sort. In addition to the Condition 1 and 2 scores, the current study was designed to include the Confirmed Correct Sorts score, which measures the number of times a subject correctly sorts the cards. This score is considered a measure of nonverbal EF in that it is based solely on correct sorts, rather than including a verbal component. Internal consistency values for Condition 1 and Condition 2 fall in the moderate to high range from .55 to .82. Due to insufficient amounts of data, these scores were not included in the final analysis.

The Tower subtest is also modified from a well-known neuropsychological test: the Tower of Hanoi (Delis et al., 2001). Higher-order abilities required to complete this task include planning, problem-solving, and inhibition. Examinees are instructed to match a target image by moving circular discs along a pegboard while adhering to a set of rules (i.e., only one disc may be moved at a time and large discs cannot be placed onto small discs). The examiner keeps track of rule violations as well as the number of moves and time taken to reach the target. A lower number of moves often indicates greater levels of planning. The Total Achievement score was utilized in the current study as an overall measure of performance across nine towers. Internal consistency values varied on this measure, ranging from .43 to .84. These values were obtained by using a Spearman-Brown formula to correct the correlation between half-tests. Test-retest reliability fell in the moderate range with a value of .51 for children and adolescents. SEM ratings ranged from 1.21 to 2.27 for this age group.

Finally, Word Context is a new test developed specifically for use in the D-KEFS to measure reasoning or problem-solving abilities (Delis et al., 2001). Subjects are presented with verbal and visual nonsense words, which are provided in sentences to help explain their meanings. Each time a sentence is provided, the examinee is allowed to guess the meaning of the word. Scores are determined by the subject’s ability to correctly define the words as well as how many sentences are provided as clues and whether the subject provides an incorrect response after answering correctly as more clues are added. The current study incorporated the Total Consecutively Correct score as a measure of EF. Internal consistency values for this score ranged from .47 to .71 in children and adolescents. A Spearman-Brown correction formula was applied to obtain split-half reliability. Test-retest reliability was rated as good with a value of .58 for this age group. SEM values were not provided for the selected score.

NEPSY, Second Edition (NEPSY-2)

The NEPSY-2 (Korkman et al., 2007) is a comprehensive neuropsychological battery for children between the ages of 3 and 16 that assesses multiple neurocognitive domains, including attention and EF, language, sensorimotor, visuospatial, memory and learning, and social perception. The NEPSY-2 was influenced by the work of Luria (1973), who was one of the first theorists to propose that multiple brain systems underlie neurocognitive functioning, particularly complex functions like EF. Additionally, his theory postulates the differing means by which brain regions interact in children versus adults; thus, the NEPSY-2 was developed specifically to identify neurocognitive dysfunction in children. Because deficits in basic processes generally underlie deficits in complex function, the NEPSY-2 is aimed at pinpointing primary dysfunction in lower brain regions to explain secondary difficulties in higher order processing (Korkman et al., 2007).

The NEPSY-2 (Korkman et al., 2007) provides numerous scores drawn from 32 subtests, which the examiner may select from and administer in any order to form an individualized battery. The test publishers provide diagnostic batteries to aid in assessment of specific conditions, including attention difficulties, language disorders, learning disabilities, and others. A general referral battery is also outlined as a recommendation for assessing a wide range of neurocognitive areas. Although the

NEPSY-2 does not provide domain or cluster scores, a variety of scores are available for interpretation. Subtest scores are reported as standard scores, scaled scores, and percentile bands. Most subtests also utilize qualitative information from behavioral observations to allow for process and error analysis. Finally, contrast scores are available as a means of comparing children's performance on different tasks. The current study utilized scaled scores from five subtests: Animal Sorting, Auditory Attention, Response Set, Inhibition, and Word List Interference. Although other NEPSY-2 subtests are proposed as measures of EF, they were not included in the current study due to age restrictions or insufficient data. Table 3.4 lists the subtests from the NEPSY-2 that were included in the analysis.

Normative data. The NEPSY-2 normative sample includes 1,200 children from the United States between the ages of 3 and 16 (Korkman et al., 2007). Stratified, random sampling was utilized in order to ensure that the sample was proportionally representative of the various ages, race/ethnicities, geographic regions, and parental education levels of the United States population according to 2003 U.S. Census data. Children with a variety of neurodevelopmental disorders were included in the sample.

Reliability and validity. Reliability coefficients, stability coefficients, and SEM values provide evidence of reliability for the NEPSY-2 (Korkman et al., 2007). Split-half procedures and alpha methods were utilized for all subtests except Auditory Attention and speeded tests, which relied upon test-retest values. Decision consistency was utilized to assess reliability of combined and contrast scores; however, the test publishers note these scores are somewhat affected by practice effects, producing lower reliabilities. The results of these studies indicate the NEPSY-2 subtests have adequate to high internal consistency with the highest coefficients obtained for verbal tests and moderate coefficients obtained for EF tests. A study of stability of performance was completed with a test-retest interval of 12 to 51 days. Pearson's product-moment correlations revealed adequate stability across age groups with slight improvement in performance observed on memory tasks as well as Inhibition, which is a measure of EF.

Table 3.4

NEPSY-2 Scores Included in the Analysis

Test – Composite Score	Area of EF Measured
NEPSY-2 – Animal Sorting	Concept Generation
NEPSY-2 – Auditory Attention	Sustained Attention
NEPSY-2 – Response Set	Set Shifting
NEPSY-2 – Inhibition (Switching)	Set Shifting
NEPSY-2 – Word List Interference	Working Memory

Note. Adapted from *NEPSY-2: A developmental neuropsychological assessment* by M. Korkman, U. Kirk, and S. Kemp. Copyright 2007 by The Psychological Corporation.

Evidence of content, construct, and concurrent validity of the NEPSY-2 has been established (Korkman et al., 2007). The test publishers argue that content validity is supported because the test relies heavily on Lurian theory and was developed and modified based on results of empirical research. Regarding construct validity, the subtests assessing similar areas of neurocognitive functioning show moderate correlations with one another while most other intercorrelations are low, which is attributed to the inclusion of children with various clinical diagnosis in the normative sample. Concurrent and convergent validity has been evaluated through the comparison of the NEPSY-2 to numerous other measures, including tests of cognitive abilities, the original NEPSY, the Children's Memory Scale (CMS; Cohen, 1997), and the D-KEFS (Delis et al., 2001). The

NEPSY-2 shows strong correlation with measures of general intelligence, such as the WISC-IV (Wechsler, 2003) Furthermore, certain domains of neurocognitive functioning correlate strongly with related domains on other assessments (Korkman et al., 2007). For example, the Attention and Executive Functioning subtests on the NEPSY-2 correlate strongly with a number of D-KEFS scores while those in the Memory and Learning domain correlate strongly with CMS scores. Performance on the NEPSY-2 was also found to negatively correlate with scores on behavior rating scales, such as the Brown Attention-Deficit Disorder Scales for Children and Adolescents (Brown ADD Scales; Brown, 2001). As ADHD symptomatology decreased, scores on the NEPSY-2 increased. Lastly, the NEPSY-2 has demonstrated differential sensitivity in detecting a variety of clinical diagnoses, including ADHD, autism spectrum disorder, learning disorders, and TBI, based on neurocognitive profiles.

Subtest descriptions. The Animal Sorting task is similar to the classic Wisconsin Card Sorting Test (WCST; Grant & Berg, 1948), which has been used in neuropsychological assessments for decades. To complete the Animal Sorting subtest, examinees must sort cards into two groups of four based on common characteristics, which are identified and selected by the examinee. (Korkman et al., 2007). This subtest assesses the child's ability to formulate novel concepts (i.e., concept formation) by identifying similarities between cards and shifting from one concept to another. The Animal Sorting Combined Scaled Score was included in data analysis for the current study as a measure of concept generation and problem-solving which are considered aspects of EF (Miller, 2013). Reliability coefficients for this subtest score were consistently high, ranging from .90 to .96. Stability coefficients and SEM values were provided for the Total Correct Sorts score with stability coefficients ranging from .63 to .73 and SEM values ranging from 1.5 to 1.8.

The Auditory Attention and Response Set subtest is divided into two major tasks, the first of which measures attention (i.e., Auditory Attention) and the second which measures EF (i.e., Response Set; Korkman et al., 2007). The first component of this subtest requires sustained attention for approximately 3 minutes as the examinee listens to an audio recording of assorted words. When a color word is read, the examinee must then quickly touch a picture of one of four colored circles. During the second component, the examinee must follow a set of rules while listening to a similar audio recording of assorted words. After hearing two specific colors, the examinee is instructed to touch the opposite of those colors. After hearing another specific color, the examinee is instructed to touch that same color. Both conditions of this task require the examinee to inhibit responses when hearing all other words than those outlined in the rules. However, the second task becomes more complex in that the examinee must hold a larger set of rules in memory and shift between those rules depending on what is heard; thus, children with EF deficits often struggle to complete Response Set with accuracy. A combined score for Auditory Attention and combined score for Response Set was utilized as two separate variables in the current study. These scores are affected by the number of errors as well as the overall completion time for each task. Reliability coefficients ranged from .71 to .91 for Auditory Attention and from .83 to .93. Stability coefficients and SEM values were not provided for either Auditory Attention or Response Set.

Inhibition is considered a measure of attention/EF (Korkman et al., 2007). Examinees must complete three successive activities or conditions: Naming, Inhibition, and Switching. The examinee must first visually attend to a series of shapes or arrows and then verbally state the names of the shapes or the direction of the arrows. Next, the examinee must inhibit his or her initial verbal responses and provide the opposite response (e.g., saying “up” when presented with a down arrow). This condition is similar to the Stroop (1935) procedure, which tests one’s ability to suppress over-learned verbal responses and provide conflicting responses (i.e., naming the color in which a color word is printed, rather than reading the color word). In the final condition, the examinee must switch between providing responses from Condition 1 and Condition 2 depending on certain rules (Korkman et al., 2007). For example, the examinee must state the initial response for symbols pictured in one color but provide the opposite response for symbols pictured in another color. The Inhibition Switching Combined Scaled score provided by the final condition of the NEPSY-2 task was utilized in the current study as a measure of shifting attention, which is considered an aspect of EF (Miller, 2013). Reliability coefficients for the final condition of this subtest are considered strong, ranging from .85 to .89. Stability coefficients for the Switching Total Completion Time ranged from .75 to .93 and SEM values ranged from 1.31 to 1.80.

Word List Interference is included in the memory and learning domain of the NEPSY-2, requiring examinees to attend to two lists of words while holding each list in memory (Korkman et al., 2007). In this way, one list serves as an interference task for the other and vice versa. This subtest is designed to measure verbal working memory. The

Word List Interference Recall score was included in the current study as a measure of EF. Reliability coefficients for this score were low, ranging from only .60 to .77. Stability coefficients were not provided and SEM values ranged from 1.44 to 1.97.

Test of Everyday Attention for Children (TEA-Ch)

The TEA-Ch (Manly et al., 1999) is a clinical battery of nine subtests designed to assess various aspects of attention, including selective attention, sustained attention, and attentional switching. The Test of Everyday Attention (TEA; Robertson, Ward, Ridgeway, & Nimmo-Smith, 1994) was originally designed to measure attention in adults but was adapted for use with children between the ages of 6 and 16. The TEA-Ch was the first assessment battery of its kind to attempt to objectively measure attentional abilities in children. Many of the tests are presented in a “game-like” format to best engage young children. Attention is considered a very difficult construct to evaluate because it can only be measured indirectly and is strongly related to other cognitive abilities. Because of these concerns, the test publishers of the TEA-Ch intentionally reduced demands of memory, motor speed, and language to more accurately measure attentional abilities.

The TEA-Ch (Manly et al., 1999) provides a variety of scores to inform practitioners during the evaluation, treatment, and management of attention problems. Additionally, two versions, A and B, are available to allow practitioners to retest children and monitor progress during or after interventions. Scores are available for each of the nine subtests, which are categorized by the type of attention each subtest measures (i.e., selective attention, attentional control/switching, and sustained attention). The subtests included in the attentional control/switching domain (i.e., Creature Count and Opposite

World) as well as those measuring divided attention and response inhibition (i.e., Walk, Don't Walk and Sky Search Dual Task [DT]) were originally intended to be included in the current study but were removed due to insufficient amounts of available data.

Normative data. The normative sample for the TEA-Ch is based on data from 293 children and adolescents between the ages of 6 and 16 years living in Melbourne, Australia (Manly et al., 1999). Stratification was used to ensure nearly equal representation within the sample of various age groups as well as gender. Children with head injury, neurological illness, developmental delays, sensory loss, attention or learning problems, and special education eligibility were excluded from the normative sample. Although normative data is somewhat limited, additional standardization and validity studies have been completed (Belloni, 2012; Chan, Wang, Ye, Leung, & Mok, 2008), which support the reliability and validity of the instrument.

Reliability and validity. Reliability information for the TEA-Ch is somewhat limited. Test-retest reliability was established by testing fifty-five children twice with an interval of 6 to 15 days between test administrations (Manly et al., 1999). Test authors parsed out age from the analysis when calculating the stability coefficients because unusually high correlations were obtained when age was not controlled. Stability coefficients, which were derived from Pearson-product moment correlations, ranged from adequate to strong for most subtests on the TEA-Ch with the lowest value observed on the Creature Counting timing score. Ceiling effects were observed on three subtests, including Walk, Don't Walk, prohibiting the use of correlation coefficients.

Regarding validity, Manly et al. (1999) argue that construct validity of the TEA-Ch is strong given that the test was developed using knowledge from research on the neurodevelopment of attention. Additionally, some measures on the TEA-Ch were adapted from well-known and valid measures of attention. For example, Code Transmission is a modified version of the Continuous Performance Test (Conners, 2013). Convergent validity is supported by studies investigating the relation of the TEA-Ch to other measures of attention, including the Stroop Color and Word Test (Trenerry, Crosson, DeBoe, & Leber, 1989), the Trails Test (Spreen & Strauss, 1991), and the Matching Familiar Figures Test (MFFT; Arizmendi, Paulsen, & Domino, 1981). The Stroop Color and Word Test correlated significantly and positively with Sky Search, Creature Counting accuracy, Map Mission, Opposite Worlds, and Code Transmission, while the Trails test correlated with Sky Search, Sky Search DT, Map Mission, Opposite Worlds, and Code Transmission. Lastly, MFFT, which is considered a measure of inhibition, correlated with Sky Search, Score, Creature Counting accuracy, Score DT, Walk, Don't Walk, Opposite Worlds, and Code Transmission.

To establish divergent validity, the TEA-Ch was compared to measures of intelligence and achievement. Small but significant correlations were obtained for Block Design and Object Assembly on the WISC-III (Wechsler, 1991) which test publishers attribute to demands on processing speed and visual-spatial thinking (Manly et al., 1999). Similarly, only measures of sustained attention correlated significantly with achievement measures on the Wide Range Achievement Test-Revised (WRAT; Justak & Wilkinson, 1984). Thus, the TEA-Ch appears to measure different constructs than intelligence and achievement measures; however, such measures do place some demands on attention, resulting in some overlap. Discriminant validity is further supported in that children with ADHD tend to exhibit significantly worse performance on all measures of the TEA-Ch except Sky Search. This same effect was not observed in a sample of children with TBI. Lastly, structural equation modeling was utilized to determine the validity of the three-factor model of attention (i.e., selective attention, sustained attention, and attentional control) utilized by the TEA-Ch. Results demonstrated that the patterns of performance observed on the TEA-Ch fit well with the three factor model.

Subtest descriptions. The Walk, Don't Walk subtest provides a measure of response inhibition (Manly, et al., 1999). On this subtest, the examinee is required to move along a pathway by touching pictures of footprints with a pen one-at-a-time after each tone heard on an audio recording. When the tone is altered slightly, the examinee must inhibit his or her automatized response by refraining from touching another footprint. Stability coefficients are not available for this subtest due to ceiling affects. Instead, percentage of agreement (71%) within one standard deviation is reported in the manual.

Creature Counting is a subtest aimed at measuring attentional control and switching (Manly et al., 1999). To complete this subtest, examinees must quickly and accurately switch between counting forwards and backwards based on visual prompts of arrows pointing up or down. In other words, the examinee must be able to shift flexibly from one mental activity to another. Stability coefficients were .71 for the accuracy score and .57 for the timing score.

The Opposite Worlds subtest also provides a measure of attentional control and switching (Manly et al., 1999). Examines must first follow a visual path by correctly naming the numbers (i.e., 1 and 2) printed along the pathway. In the next portion of the task, examinees must follow a similar path but say *two* when presented with the number 1 and say *one* when presented with the number 2. The stability coefficient for this subtest was .85.

The Sky Search DT subtest evaluates an examinee's ability to divide his or her attention (Manly et al., 1999). This subtest requires the examinee to combine two previously completed tasks by searching for pairs of identical spaceships while counting a series of sounds. This dual task requires the use of EF systems to accurately maintain attention to both activities while simultaneously completing them with efficiency. The stability coefficient for this subtest was .81.

Woodcock Johnson Tests of Cognitive Abilities, Third Edition (WJ III COG NU)

The Woodcock Johnson Tests of Cognitive Abilities, Third Edition Normative Update (WJ III COG NU; McGrew et al., 2007) is a test battery that provides a comprehensive measure of intellectual abilities encompassing a large age range (ages 2 to 95 years). The test battery is derived from the Cattell-Horn-Carroll (CHC) theory of intelligence (McGrew et al., 2007). This theory proposes that cognitive abilities are organized in a hierarchical manner with 70 narrow abilities falling under nine factors of intelligence or broad abilities. This framework was developed from Carroll's three stratum theory of intelligence (1997) as well as the older Gf-Gc theory of intelligence (Horn & Cattell, 1966).

The WJ III COG NU includes 31 subtests, which are divided into three batteries: standard, extended, and diagnostic (McGrew et al., 2007). Scores provided by the test include a general intellectual ability (GIA) score as well as various composite scores including those assessing seven broad abilities (i.e., Comprehension-Knowledge, Long-Term Storage and Retrieval, Visual-Spatial Thinking, Auditory Processing, Fluid Reasoning, Processing Speed, and Short-Term Memory). Each broad composite score is drawn from performance on two subtests assessing different narrow abilities. Other composite scores assess other cognitive constructs, including broad attention, working memory, selective attention, and executive processes. Because these cognitive processes have been proposed as subcomponents of EF (Barkley, 2012b; McCloskey & Perkins, 2012), the subtests comprising these composite scores were selected for inclusion in the analysis for the current study as well as those assessing higher-order thinking (i.e., problem-solving). However, only Analysis/Synthesis, Concept Formation, and Numbers Reversed were utilized in the analysis. Auditory Attention, Auditory Working Memory, and Planning were excluded due to insufficient available data. Table 3.5 lists the subtests from the WJ III COG NU included in the analysis.

Table 3.5

WJ III COG NU Scores Included in the Analysis

Test – Composite Score	Area of EF Measured
WJ III COG NU – Analysis/Synthesis	Deductive Reasoning
WJ III COG NU – Concept Formation	Inductive Reasoning
WJ III COG NU – Numbers Reversed	Working Memory

Note. Adapted from *Woodcock-Johnson III Technical Manual* by K. S. McGrew and R. W. Woodcock. Copyright 2001 by Riverside Publishing

Normative data. Normative data was drawn from a sample of 8,818 individuals between the ages of 2 and 90 from over 100 geographic regions (McGrew et al., 2007). The school-age sample, which was utilized in the current study, was composed of 4,783 children. Efforts were made to obtain a normative group representative of the U.S. population through a three-stage sampling process, which used random sampling and a stratified sample design. The following variables were controlled for: geographic region, community size, sex, race/ethnicity, type of school (i.e., public, private, or home) and parent education and occupational status.

Reliability and validity. The WJ III COG demonstrates strong reliability as evidenced through internal consistency ratings, test-retest correlations, inter-rater reliability, and SEM values (McGrew et al., 2007). Split-half procedures were utilized to determine test reliabilities by comparing odd and even test items and using a Spearman-Brown correction formula. Rasch analysis procedures were incorporated for speeded tests without odd and even items, such as Planning. Composite scores showed very strong reliability coefficients, which were generally in the .90s range. The individual subtests to be included in the current study had internal consistency ratings of .75 and above.

In regards to validity, the WJ III COG NU is considered an empirically sound instrument due to strong evidence of its validity drawn from previous research, expert opinion, and factor analyses (McGrew et al., 2007). Content validity is considered strong because the battery is founded on a well-established theory of intelligence (i.e., CHC theory). Subtests were designed in a way that allows each to assess a different narrow ability contributing to a broad ability. Construct validity is demonstrated through confirmatory-factor analytic models, which support the use of a seven-factor model of intelligence. Cognitive cluster correlations are low, supporting the existence of separate domains of intelligence. Furthermore, criterion validity is strong in that scores obtained on the WJ III COG NU correlate strongly (*r* = .70 or greater) with other measures of intelligence, including the Wechsler Preschool and Primary Scales of Intelligence-Revised (WPPSI-R; Wechsler, 1989), Wechsler Intelligence Scale for Children-Third Edition (WISC-III; Wechsler, 1991), Differential Ability Scales (DAS; Elliott, 1990), Kaufman Adolescent and Adult Intelligence Test (KAIT; Kaufman & Kaufman, 1993), and Stanford-Binet Intelligence Scale-Fourth Edition (SB-IV; Thorndike, Hagen, & Sattler, 1986). The WJ III COG NU also exhibits predictive validity in its ability to discriminate between children with and without learning disabilities, ADHD, anxiety, head injury, and language disorders (McGrew et al., 2007).

Subtest descriptions. The Auditory Attention subtest measures one's ability to discriminate sounds when presented with distracting auditory stimuli (McGrew et al., 2007). The examinee listens to a word and then touches one of four pictures to indicate what he or she heard. As the test progresses, the distracting noises grow louder. This subtest was included as a measure of EF because it requires the ability to selectively attend to specific stimuli while suppressing irrelevant stimuli. Reliability coefficients ranged from .83 to .93 for the 5 to 19 years of age group. SEM values for standard scores ranged from 3.97 to 6.18.

Analysis/Synthesis measures the ability to engage in sequential reasoning, an aspect of fluid reasoning that involves drawing conclusions from provided information

(McGrew et al., 2007). To complete this task, examinees must perform increasingly complex procedures, similar to a mathematical system, using instructions provided symbolically. Corrective feedback is provided during the first portion of the task to refine the examinee's ability to complete each procedure. Reliability coefficients ranged from .81 to .94 for the 5 to 19 years of age group. SEM values for standard scores ranged from 3.61 to 6.47.

The Concept Formation subtest is a controlled-learning task designed to measure the ability to utilize induction, an aspect of fluid reasoning (McGrew et al., 2007). Examinees are presented with a set of visual stimuli from which they must determine a rule for the set. The examiner provides corrective feedback for incorrect items, excluding those at the end of the subtest. Test items also must be completed within a certain time frame. The test authors propose that, in addition to induction, the test assesses EF in that it requires examinees to flexibly shift between mental sets. Reliability coefficients ranged from .93 to .96 for the 5 to 19 years of age group. SEM values for standard scores ranged from 3.04 to 3.94.

Auditory Working Memory measures short-term and working memory stores and divided attention (McGrew et al., 2007). To complete this subtest, an examinee must attend to a series of numbers and words, mentally reorder the information, and list the words in sequential order followed by the numbers in sequential order. Because this task requires the examinee to shift attentional resources between two groups of information, it is also viewed as a measure of EF capability. Reliability coefficients ranged from .80 to

.93 for the 5 to 19 years of age group. SEM values for standard scores ranged from 4.11 to 6.79.

Numbers Reversed assesses short-term and working memory stores by measuring the span of numbers an examinee can hold in his or her immediate awareness and then verbally list backwards (McGrew et al., 2007). Certain theorists consider working memory a component of EF (Barkley, 2012b; McCloskey & Perkins, 2012). Reliability coefficients ranged from .84 to .92. SEM values for standard scores ranged from 4.24 to 6.03.

Finally, the Planning subtest is considered a measure of EF in that it assesses one's ability to identify and apply solutions in order to trace a pattern along a dotted line without lifting the pencil or retracing (McGrew et al., 2007). Additional skills in fluid reasoning and visual processing are required to successfully complete this task. Reliability coefficients ranged from .63 to .78 for the 5 to 19 years of age group. SEM values for standard scores ranged from 7.04 to 9.12.

Data Analysis

The primary purpose of this study was to determine the relationship between cognitive and behavioral measures of EF in a mixed clinical sample through three questions. These questions involved the comparison of scores obtained on cognitive measures of EF and parent ratings of behavioral EF. Using the BRIEF and the BASC-2 as measures of behavioral EF as well as a variety of cognitive tasks as measures of cognitive EF in a clinical group the following questions were posed:

1. What factor structures are obtained for both behavioral EF (i.e., parent ratings) and cognitive EF (i.e., performance-based measures)?
2. Are the factor structures obtained for behavioral EF and cognitive EF comparable? In other words, do the behavioral EF factors correlate with the cognitive EF factors?
3. Finally, do the subtests loading on behavioral EF factors predict those loading on correlated cognitive EF factors?

In regard to the first research question, it was anticipated that multiple factors would be obtained for behavioral and cognitive EF measures. The Integrated SNP/CHC model (Miller, 2013) suggests that four domains of EF exist: cognitive flexibility, concept recognition and generation, problem solving, fluid reasoning, and planning, and response inhibition. Although attention and working memory are classified as cognitive facilitators/inhibitors in Miller's model (2013), other researchers (Barkley, 2012b; McCloskey & Perkins, 2012) suggest these cognitive abilities are subtypes of EF. Therefore, those abilities were expected to load onto separate factors during the EFA stage of the data analysis. It was anticipated that the behavioral measures of EF would load onto one behavioral and one cognitive factor based on factor analyses previously conducted using the BRIEF (Gioia et al., 2000).

Second, it was anticipated that there would be some similarity between the factor loadings obtained for the cognitive and behavioral EFAs. This hypothesis was based upon research showing that scores on the BRIEF correlate strongly with scores assessing related domains on the BASC and with classical measures of EF (Reynolds & Kamphaus, 2004). Additionally, Korkman et al. (2007) found that as performance on the NEPSY-2 improved, ADHD symptomatology, as rated on the Brown ADD Scales (Brown, 2001), decreased. Together, these findings suggest some degree of a relationship exists between behavioral and cognitive measures of EF. However, it was anticipated that there would be some degree of difference between the factors obtained on the two EFAs given research revealing low correlations between scores on EF tasks and behavioral ratings by caregivers (Wilson, 1998). This study was designed to help clarify the degree to which cognitive and behavioral measures of EF can be compared.

In regards to the final research question, it was anticipated that the relationship between cognitive and behavioral measures of EF would be strong enough that parent ratings of certain EF skills would predict children's performance on correlated cognitive measures of EF. This was based on the idea that EF strongly influences behaviors and that behaviors serve as physical manifestations of inner functioning. For example, those with EF dysfunction, including children with ADHD and autism spectrum disorder, commonly exhibit specific behaviors, such as impulsivity, that are associated with that underlying dysfunction (Barkley, 2012b). In this way, it was expected that certain types of behaviors related to EF dysfunction, as observed by parents and caregivers, would predict poor performance on related EF scales. For example, a child with clinically significant ratings of Inhibition on the BRIEF or Hyperactivity on the BASC-2 might exhibit poor performance on Color Word Interference 4, which is purported to measure Inhibition.

Descriptive Statistics

Prior to examining the research questions, descriptive statistics of the sample and of all scores were calculated using version 19 of the Statistical Package for the Social Sciences (SPSS). Demographic information was examined, including the number of participants classified by gender, disability, and race/ethnicity in order to indicate the degree to which the results can be generalized to the larger population of children with clinical diagnoses. Scores for behavioral and cognitive measures of EF included *T*-scores, standard scores, and scaled scores. The *T*-scores were derived from behavioral measures of EF while the standard and scaled scores were derived from cognitive measures of EF. Distributions were also analyzed to identify the degree of normality and linearity of each cognitive and behavioral EF score in the dataset and ensure that all assumptions were met; however, as indicated in the manuals for the BRIEF and BASC-2, these scores are commonly positively skewed. Extreme outliers were identified but were not excluded from the analysis because they were not found to significantly affect results.

Bivariate Correlations

Bivariate correlations between each variable (i.e., composite scores on the BRIEF and BASC-2 and subtest scores on the D-KEFS, NEPSY-2, TEA-Ch, and WJ III COG NU) were completed using SPSS. A correlation matrix was generated, providing correlation coefficients, significance levels, means, and standard deviations for each pair of variables. Pearson product-moment correlation was utilized to indicate the degree to which variables were related to one another based on how much variance was shared by those two variables (Meyers et al., 2006). Correlation values (*r)* range from -1.0 to 1.0

with -1.0 indicating a perfect negative relationship (i.e., as one value increases by one unit, the other decreases by one unit) and 1.0 indicating a perfect positive relationship (i.e., as one value increases by one unit, the other also increases by one unit). Cohen (1988) suggests values of .5, .3., and .1 indicate large, moderate, and small correlations respectively. Scatterplots were used to visualize the correlation matrix. When two variables are placed on separate axes, the data points may cluster together in an oval pattern. If the pattern moves from left to right in an upward fashion, the relationship between variables is positive. If the pattern moves from left to right in a downward fashion, the relationship between variables is negative. Multicollinearity was considered before moving forward with analyses. Although many behavioral variables were correlated, none were found to be very strongly correlated (i.e., *r* values greater than 0.9) or perfectly correlated (i.e., singularity; Field, 2009). Therefore, multicollinearity, which occurs when two scores are so highly correlated that the inclusion of both scores in a statistical analysis has the potential to reduce the statistical power of the analysis, was not expected to be negatively affect the results of this study (Meyers et al., 2006).

Exploratory Factor Analysis

Exploratory factor analysis (EFA) is considered a data reduction method whereby factors with similar characteristics are extracted from large sets of data by identifying the variables that are highly correlated with one another and independent of other variables (Meyer et al., 2006). Although factors are comprised of variables, factors are considered latent variables because they are not directly measured but are instead assumed to exist based on the variables that have been directly measured. EFA is often used in test development and evaluation in order to identify the subtests that assess certain cognitive processes. Spearman's (1904) two-factor theory of intelligence is thought to have originated the idea of investigating factor structures. When conducting factor analysis, larger sample sizes are considered best. Comrey and Lee (1992) suggest guidelines for sample sizes when conducting factor analysis. A sample size of 50 cases is considered poor, whereas a sample of 200 is viewed as fair, and a sample of 1,000 is considered excellent. Meyers et al. (2006) recommend 10 participants per variable with a sample size no smaller than 200. The three key steps in factor analysis include the following: the generation of a correlation matrix, the extraction of the factors, and rotation.

In EFA, variables that are strongly related in either direction are generally indicators for a factor (Meyers et al., 2006). When variables are strong indicators of a factor, they are assigned larger weights, whereas when variables are weak indicators of a factor, they are assigned smaller weights. Prior to conducting the factor extraction stage, the Kaiser-Meyer-Olkin Measure of Sampling Adequacy (MSA) and Bartlett's test of sphericity should also be completed in order to determine whether the data is worth reducing (Field, 2009). The Kaiser-Meyer-Olkin MSA is used to determine the amount of variance in variables that can be explained by underlying factors. The statistic is calculated by comparing the magnitudes of the correlation coefficients to those of partial correlation coefficients, which should be small if two variables share a factor. High values (i.e., those close to 1.0) suggest that EFA may be favorable; however, values less than 0.5 suggest that EFA will not be useful. Bartlett's test of sphericity can also be used to test the hypothesis that the correlation matrix is an identity matrix, indicating that the variables are not strongly related enough to accurately detect a factor structure.

The next step in EFA is to extract the factors (Meyers et al., 2006). This stage is aimed at accounting for all variance. As one factor is extracted it accounts for a proportion of total variance. Each successive component accounts for another portion of variance until all original variance is accounted for. In other words, when combined, all variables should equal the amount of total variance. The simplest form of EFA is considered to be principal components analysis (PCA). In this analysis, a component can be defined as a weighted linear combination pictured as a straight line that combines all variables being analyzed. Variables are weighted along the line based on the degree to which they contribute to the principal component. The number of variables is equal to the number of components in the analysis.

The extraction phase of PCA can be visualized as a three-dimensional cube with various data points representing variables contained inside the cube (Meyers et al., 2006). Points that are closer together are more strongly correlated than those that are further apart from one another. In this way, distance between variables indicates correlation. During extraction, lines representing the principal component are fitted into the cube one component at a time. Before lines are fitted, no variance is accounted for, but as lines are added, they account for greater levels of variance until all variance is accounted for. Lines of best fit account for the greatest amount of variance and satisfy the least squares rule. They are placed into the cube space in a way that comes as close as possible to all the variables. This one location, or linear function, allows for the smallest sum of squares calculated by the distances of the variables from the line (i.e., least squares rule). Sums of squares are determined by r^2 values, which measure the distance between variables and components. After the first line is extracted, or fitted, a second component is extracted and must intersect the first line at a 90-degree angle. In this way, the components are orthogonal to each other, meaning they are uncorrelated. As additional lines are subsequently fitted, the components intersect each other perpendicularly and are placed in the locations that account for the greatest level of variance remaining.

Although components can continue to be extracted until all variables have been included, it is often not necessary to arrive at this complete solution because some of the variables account for very little variance (Meyers et al., 2006). Therefore, the extraction phase can be ended when enough components have been extracted to explain a large degree of variance. Generally, factors with eigenvalues below 1 are not included in the solution. Eigenvalues provide an overall measure of each component. The eigenvalue is calculated by adding all of the r^2 values along the line. The most optimal eigenvalues are maximal because they represent a position where the squared distances from the line are lowest, resulting in stronger correlations and higher r^2 values. Eigenvalues can be interpreted as demonstrating the units of variance each factor accounts for. For example, an eigenvalue of 3.25672 for one factor would suggest that factor accounted for 3.25 units of variance. If this value is divided by the total amount of variance, the percent of total variance extracted by the factor can be determined.

Principal axis factoring (PAF) is one of many procedures that may be utilized when conducting exploratory factor analysis and is slightly more complex than PCA

(Meyers et al., 2006). The current study utilized the principal axis extraction method because the goal of the study was to explore the theoretical factor structure of test instruments rather than reduce data (e.g., create composite scores from subtest scores). This method differs from PCA in that it uses communalities (i.e., a measure of shared variance) on the diagonal of the correlation matrix rather than ones. This allows for the assumption that measurement error exists, which is important when utilizing measurement tools with some degree of error, such as those measuring EF. In order to utilize communalities, squared multiple correlations are calculated to indicate the relationship between each variable and the remaining set of variables. Low communalities indicate that the variables analyzed have little in common and thus cannot reliably predict a factor solution. After communalities are estimated, the EFA is conducted in the same manner as PCA. PAF was selected for use in the current study because it is considered a conservative means of estimating the variables' reliability. It is less likely than PCA to provide an inflated solution that accounts for more variance. Because EF scales and assessments do not directly measure EF, it is important to utilize an extraction method that accounts for error.

Following extraction, the factor rotation phase is completed. During this phase, the extracted factors are pivoted around the location at which they intersect (Meyer et al., 2006). This phase is important because it improves the ability to interpret the factors (Field, 2009). This is accomplished by maximizing the loading of each variable on one factor while simultaneously minimizing the loading on the other factors through redistribution of variance. The concept of rotation was first introduced by Thurstone

(1947) and then adapted by other researchers in the 1950s into a procedure known as quartimax. However, these methods produced poor outcomes and were often difficult to apply to datasets (Meyers et al., 2006). Kaiser (1958) introduced a varimax strategy, which is the rotation method generally preferred by researchers today.

Ultimately, the goal of rotation is to achieve simple structure (Meyers et al., 2006). This concept is best understood by returning to the idea of the least squares rule. Because lines are fitted in a way that reduces the distance between the factor and the variables, most of the variables do not end up perfectly positioned on the line. By rotating all of the factors together but leaving the variables in place, the factors are more likely to closely align with some of the variables. In other words, the correlations between the variables and factors should be very high (i.e., near 1) or very low (i.e., near 0) when simple structure has been achieved. In this way, the factor moves closer to some variables but farther away from others. During varimax rotation, the factors remain orthogonal to each other before, during, and after the rotation process. The total amount of variance accounted for should remain unchanged following rotation.

In contrast to orthogonal rotation, oblique rotation does not require the factors to be uncorrelated (Meyers et al., 2006). Rather than pivoting the factors at a perfect 90-degree angle, the factors may be pivoted at an oblique angle as if a hinge connects them at their point of intersection. The primary oblique procedures available are Direct Oblimin and Promax. Direct Oblimin is generally preferred but Promax can be utilized when analyzing very large data sets. The Direct Oblimin method allows the researcher to control the amount of correlation between factors by setting a maximum value permitted between factors. This value is important because if two factors are allowed to strongly correlate they may be too similar and thus may not indicate the existence of two separate factors.

Two separate factor analyses were conducted as part of the current study: one investigating the factor structure of cognitive measures of EF and another investigating the factor structure of behavioral measures of EF. In order to determine the most appropriate procedure for use in the current study, correlations were analyzed according to the guidelines published by Meyers et al. (2006). Because many correlations between variables were valued at .3 or higher, oblique rotation was utilized (i.e., Direct Oblimin).

Several components must be analyzed when interpreting the results of EFA. First, the scree distribution, which plots eigenvalues should be examined to determine the number of factors to include in the solution (Field, 2009; Meyers et al., 2006). As mentioned previously, factors with eigenvalues below one are generally not included in the solution. On the scree plot, factors that show a reasonable drop from the line are generally included because they continue to contribute to variance. However, once the factors begin straightening out on the distribution, they do not account for a meaningful level of variance. The location at which the line begins to straighten is called the point of diminishing returns because there is no benefit in adding additional factors.

After analyzing eigenvalues on a scree plot, researchers interpret the magnitude and pattern of coefficients or loadings on the rotated factor matrix (Meyers et al., 2006). Comrey and Lee (1992) categorize coefficients in order to indicate the degree to which variables relate to each factor. These categories are outlined as follows: .70 is excellent,

.63 is very good, .55 is good, .45 is fair, and .32 is minimal. Most researchers do not accept values below .30 but .40 is viewed as more conservative value, particularly when many variables are included in the analysis and the sample size is low (Stevens, 2002). When these values are met, variables may be considered as related to or “loading” on a factor (Meyers et al., 2006). Factors can be interpreted by identifying similarities between the variables loading on one factor. For example, in the current study, while interpreting the results of an EFA exploring cognitive measures of EF, similarities between subtests loading on each factor were identified and outlined. Factors were then be named based on the similarities between the variables loading on each one.

Correlation Between Factors

The results of EFA are considered appropriate in allowing researchers to compare factor similarity (Ferrando & Lorenzo, 2000). This step will allow the factors obtained from results of cognitive and behavioral EF testing to be compared. Initially, Tucker’s coefficient of congruence was identified as a means of comparing factor structures. The concept of comparing factor structures was first introduced by Burt (1948) and originally termed unadjusted correlation. It was then renamed congruence coefficient by Tucker (1951). As with bivariate correlations, *r* values obtained range from -1 to +1 with scores of -1 and +1 indicating perfect negative and positive correlations, respectively (Lorenzo-Seva & ten Berge, 2006; Abdi, 2007; Wuensch, 2016). The congruence coefficient differs from Pearson’s *r* because it is computed from the deviation of factor loadings from zero, rather than the deviations from the means of the factor loadings (Jensen, 1998). Because of this, the congruence coefficient is less likely to produce misleading results than

Pearson's *r.* Generally, congruence coefficient values in the range of .85 and .94 indicate factors are fairly similar while values above .95 suggest the factors should be considered equal (Lorenzo-Seva & ten Berge, 2006).

Unfortunately, a congruence coefficient was not able to be calculated during the current study because the standard deviations of the scores being compared differed. Therefore, in order to answer the second research question, a series of simple Pearson product moment correlations were conducted using mean scores of all factors. The goal of this analysis was to determine the relationship between factors extracted during each of the factor analyses. Scores for each factor were calculated by finding the mean of the variables making up the factors for each participant. Mean scores were not calculated if cases did not have data for the majority of the variables making up the factors.

Linear Regression

The final step in the data analysis was to conduct a series of linear regression analyses to determine whether the scores loading on each behavioral EF factor predicted those loading on similar cognitive EF factors. Regression is utilized to move beyond correlation by determining whether the values of one variable can be predicted by the values of another variable (Meyers et al., 2006). Simple linear regression is utilized when one variable predicts another single variable. In regression, a line is fitted to the scatterplot and oriented in the same direction as the oval formed by the data points. The regression line is used to predict the variable on the *Y* axis based on the values of the variable on the *X* axis. Similar to factor analysis, the regression line is found by adhering to the least squares procedure. In other words, the line is fitted in a location where the sum of squared distances is lowest.

Regression equations, which represent the predictive relationship between two variables, are formed for both raw data and standardized data (i.e., *Z* scores; Meyers et al., 2006). In the raw equation, a b weight or b coefficient, which is the slope of the regression line, is used to indicate the amount of change that can be predicted in the variable on the *Y* axis (i.e., the criterion variable) based on a certain amount of change in the variable on the *X* axis (i.e., the predictor variable). In the standardized equation, raw scores are converted to *Z* scores, which have a mean of 0 and a standard deviation of 1. Beta weights or beta coefficients are used in place of b weights or b coefficients.

Multiple regression analysis is utilized to predict a linear relationship between multiple variables (Field, 2009). The goal of multiple regression is to create a model or linear equation that indicates the combination of independent variables that best predicts an dependent or criterion variable (Meyer et al., 2006). As with simple regression, the regression line is fitted into a location that adheres to the least squares rule. This method reduces residuals, or the data points that fall away from the line. The regression line represents predicted values. Its' slope and intercept allow for the creation of a regression equation that can be used to predict the outcome or criterion variable by plugging in available data points in the predictor variables. Because multiple factors contribute to behavior and cognition in human subjects, the inclusion of more variables in the equation often allows for more accurate predictions.

The standard regression method, which may also be referred to as the simultaneous or direct method, is utilized when researchers wish to leave the regression method unspecified (Meyers et al., 2006). This method allows all predictors to be entered into the solution in a single step. Although they are entered at the same time, predictors are individually evaluated in that they are weighted as if they were entered into the equation after the other variables. Each predictor must account for separate portions of the total variance in the criterion variable. Therefore, when other variables already account for a certain amount of variance, the remaining variables target residual variance, which is the remaining amount of variance after the effects of the other variables have been removed or partialled out. Partial correlation is used to analyze the relationship between variables when other variables have been controlled for. After the b and beta weights have been calculated for one predictor variable, the process is repeated to analyze the remaining variables and determine their b and beta weights. The current study utilized the standard method because the predictor variables were selected based on the results of EFA. Because the predictors entered into each model loaded on the same factor they were considered important enough to be included in the model even if they do not contribute significantly to the model.

The final regression model demonstrates that the predictor variables explain a certain amount of variance in the outcome or criterion variable (Meyers et al., 2006). Squared semipartial correlations can be analyzed to determine how well the model works in explaining the predictor variable. These correlations allow the researcher to determine the unique contribution of each predictor. Larger numbers indicate that the associate variables make a larger contribution to the criterion variable. However, when predictor variables are strongly correlated, they may not explain a significant amount of variance on their own.

When interpreting the results of a regression model, various values can be examined (Meyers et al., 2006). Because numerous variables are involved in regression, a simple Pearson correlation coefficient (*r*) cannot be used. Instead, a multiple correlation coefficient is used (*R*). The squared multiple correlation or coefficient of multiple determination (R^2) indicates the degree to which one variable is associated with other variables. Some researchers prefer to interpret the adjusted R^2 value because it reduces the amount of inflation produced by measurement error. However, this is most important when analyzing data with small sample sizes (i.e., less than 20 cases per predictor). Because the sample size of the current study was 176, the R^2 value was interpreted in a similar fashion as *r* values obtained during correlation analyses. Additionally, b and beta weights were interpreted. b weights indicate the amount of change in the predictor variable that predicts one unit of change in the criterion variable. For example, a b weight of 2.75 would suggest that an increase of 2.75 in the predictor variable would predict an increase of 1 in the outcome variable. Because beta weights are in a standardized *Z*-score form, they allow researchers to compare the contribution of one variable to that of other variables.

An important component of regression analysis is the variate, which is the weighted combination of variables. The variate is often viewed as representing an underlying construct or latent variable that is used to predict the criterion variable. In the context of the current study, these variates were considered different EF factors determined through EFA. Behavioral scores loading on a factor were used to together predict performance on related cognitive scores; thus, the variates in the current study were considered factors or subtypes of EF. Regression equations were used to indicate whether certain latent variables, representing EF subtypes, can be used to predict scores on measures of cognitive EF. This phase of the data analysis was designed to provide information regarding the ability to make predictions about how a student will perform on cognitive measures based on behavioral ratings completed by their parents. In other words, if behavioral ratings of EF were found to predict how a child performs on cognitive measures of EF, these results would suggest behavioral ratings might be a valid means of assessing a child's EF abilities.

Summary

The current was designed determine the relationship between parent ratings of EF and performance on cognitive measures of EF in a mixed clinical sample through the use of exploratory factor analysis, correlation, and regression analyses. These methods were selected as a means of indicating whether behavioral ratings of EF have a similar factor structure as cognitive or performance-based measures of EF. The literature review from the previous chapter provides evidence that the definition and structure of EF remains in contention. Therefore, this study was aimed at attempting to clarify the components of EF measured by both behavioral rating scales of EF and performance-based EF tasks and then comparing the structures of the two types of measures. Additionally, it was anticipated that results would inform assessment practices by providing information as to whether behavioral scales may be used as an appropriate means of assessing a child's EF abilities in place of time-consuming neuropsychological assessments.

CHAPTER IV

RESULTS

This study was designed to explore the factor structure of both performance-based or cognitive measures of EF and behavioral measures of EF (i.e., parent ratings) and to compare the factor structures obtained for each type of measure. It was anticipated that distinct factor structures would be observed for both cognitive and behavioral measures and that behavioral measures loading on specific factors would predict cognitive measures loading on similar factors. To investigate the above research questions, exploratory factor analysis (EFA), correlation, and linear regression analyses were completed. The results of these analyses are delineated in this chapter.

Missing Data

Prior to conducting the primary analyses, a missing data analysis was completed. Missing data have the potential to negatively impact the validity of a study and therefore should be examined prior to beginning a statistical analysis (Little & Rubin, 2002). In order to reduce the amount of missing data included in the current analysis, specific guidelines were followed while selecting cases and variables for inclusion. Specifically, variables were excluded from the analysis if over two-thirds of the participants had missing data with the exception of Activities of Daily Living, which was included to ensure all BASC-2 variables were utilized. The TEA-Ch variables were completely excluded from the analysis because over two-thirds of all participants were missing data for each subtest. Additionally, Auditory Attention, Auditory Working Memory, and Planning from the WJ III COG NU and Sorting (Condition 1, 2, and Confirmed Correct) from the D-KEFS were excluded. Therefore, 15 cognitive variables, rather than the initially proposed 25 variables were used as part of the current study. Cases were also excluded from the analysis based on the following criteria: missing all data for either the BASC-2 or BRIEF and/or missing data for over two-thirds of the cognitive variables. Using these criteria, the final dataset included 22 behavioral variables and 15 cognitive variables with 176 cases, each of which contained data from the BASC-2, BRIEF, and at least one third of the cognitive variables.

A number of practices are available to account for missing data but decisions should be based largely on the characteristics of missingness. Large samples, are generally less affected by missing data than those with small sample sizes. If incomplete values are above 5%, researchers recommend treating the data in some way (Little & Rubin, 2002; Schafer, 1999). Additionally, the pattern of data is important to consider. If data are missing completely at random (MCAR), data may be imputed without significant validity concerns.

Results from the missing value analyses using SPSS indicated that out of 176 recorded cases, all cases contained some missing data (100%) and out of 37 variables, 30 variables contained missing data (81.08%) which amounted to a total of 23.42% missing information in the dataset. To assess whether the pattern of missing values was missing completely at random (MCAR), Little's MCAR test (Little, 1988) was conducted. The null hypothesis of Little's MCAR test is that the pattern of the data is MCAR and follows a $\chi 2$ distribution. Using an expectation-maximization algorithm, the MCAR test estimates the univariate means and correlations for each of the variables. The results revealed that the pattern of missing values in the data was MCAR, $\chi 2$ (4841) = 4834.572, p = .523. In order to address the missing data, full information maximum likelihood (FIML) and pairwise deletion were utilized.

FIML and pairwise deletion are best used with data that are MCAR (Collins et al., 2001). FIML works by estimating the parameters of a statistical model by analyzing available data. Parameter values that maximize the likelihood of obtaining the available data are included in the model. This procedure was used during the exploratory factor analysis stage. Pairwise deletion was utilized during the correlation and regression stages of the analysis. This method attempts to minimize loss of data by allowing cases containing some missing data to be included in the analyses (Kang, 2013). Cases are excluded from analyses utilizing variables with missing data but included when they contain the necessary information.

Basic Assumptions

Characteristics of the data, including sample size, outliers, normality, and multicollinearity, were investigated in order to determine that basic assumptions were met prior to conducting the analyses. A variety of statistical analyses are available to evaluate the adequacy of sample sizes; however, there is no method to determine statistical power when conducting EFA but multiple researchers have provided guidelines. Nunnally (1978) states that there should be at least 10 participants per variable in order to reduce sample error. Gorsuch (1983) recommended sample sizes of at least 100. Comrey and Lee (1992) provide more specific recommendations, describing sample sizes of 100 as poor, 200 as fair, 300 as good, 500 as very good, and 1000 as excellent. Sample size is less important when communalities obtained during the factor analysis are greater than .6 (MacCallum, Hong, Widaman, & Zhang, 1999).

Following the above guidelines, the sample size of 176 is considered adequate for both analyses. Both the behavioral and cognitive EFAs adhere to the recommendations of Gorsuch (1983) and the sample size would be considered fair by Comrey and Lee (1992). While the cognitive EFA meets the guideline of Nunally (1978), the behavioral EFA falls slightly below this guideline, given that 22 variables were utilized in the analysis. Therefore, communalities were carefully investigated as recommended by MacCallum et al. (1999) in order to determine interpretability.

Outliers, which are known as extreme values, can bias parametric tests. They are defined as values three times the interquartile range beyond the 25^{th} or 75^{th} percentiles (Field, 2009). Six extreme outliers were detected using SPSS. These outliers were observed in the Daily Living and Conduct Problem variables of the BASC-2 and the Concept Formation variable of the WJ III COG NU. Three of these outliers were identified as administrative errors and were recoded as missing data. Analyses were conducted with and without the other three extreme outliers. Significant differences were not noted during comparison of the findings. Therefore, data reported here were drawn from the analyses conducted with the three extreme outliers included.

Most statistical analyses assume that data are drawn from a population that follows a normal distribution; however, most data vary slightly from normality (Meyers et al., 2008). Normality tests, including Shapiro-Wilks and Kolmogrov-Smirnov tests can be completed for sample sizes smaller than 100. However, larger sample sizes are less impacted by skewness and kurtosis. To investigate the distribution of the data in the current study, histograms, Q-Q plots, and box plots were created. All data exhibited a normal distribution with the exception of certain BRIEF and BASC-2 variables, which were slightly skewed right. The skewness was not considered significant enough to utilize a log transformation, particularly because the size of the sample exceeded 100 (N = 176). Additionally, the handbooks for these assessments indicate that scores tend to be positively skewed. Therefore, the distribution of the data was considered adequate for completing the analyses.

Data were also analyzed for multicollinearity while running the series of regression analyses. Multicollinearity occurs when predictor variables are so strongly correlated that they are considered near perfect linear combinations of each other (Field, 2013). Multicollinearity poses an issue in regression analysis because it increases the likelihood that the coefficient estimates and standard errors will become inflated. Regression analyses were completed in Stata, rather than SPSS, in order to obtain a variance inflation factor (VIF), which is computed for each variable when commanded. VIF scores falling above ten warrant further investigation, while scores falling below that number indicate that multicollinearity is not an issue. All VIF scores in the current study fell below 5, indicating that no variables were perfect linear combinations of one another and therefore, multicollinearity did not impact the results.

Descriptive Statistics

The final sample consisted of 176 male and female participants between the ages of 8 and 16 years with a variety of clinical diagnoses. The average age for the sample was 11.11 years (SD = 2.51). The majority of the participants were male (61.9%), which is likely because school-age males are more frequently given clinical diagnoses, including ADHD, than school-age females (Gershon, 2002). Ethnicity information was reported for 172 out of 176 participants with 119 identifying as Caucasian, 28 as Hispanic/Latinx, 15 as African American, 2 as Pacific Islander, and 8 as "other." All 176 participants were given a broad diagnosis, including attention deficit/hyperactivity disorder (*n* = 61), learning disability (*n* = 53), autism spectrum disorder (*n* = 14), general medical diagnosis (n = 13), emotional disability (n = 12), neurological impairment (*n* = 9), intellectual disability (*n* = 4), and other/multiple disabilities (*n* = 10).

Additionally, intelligence quotient (IQ) scores were examined. Overall IQ scores were provided for 112 of 176 participants and were drawn from performance on a variety of norm-referenced assessment tools (WJ III COG NU, WJ IV COG, KABC-2, WISC-IV, WISC-V, and SB-5). The average IQ score for the sample was 91.80 (SD = 14.75). Scores ranged from 48 to 129 with eleven scores falling below 70. The mean score of this sample is slightly below but within one standard deviation of the average score of a general population based on the normal curve (Field, 2009). This was to be expected because the sample was made up of children with clinical diagnoses. The same pattern was observed when analyzing performance on the behavioral and cognitive measures of EF. In other words, participants' behavior was rated as more severe than average and performance fell slightly below average. More detailed descriptive statistics for each cognitive and behavioral variable, including the number of missing cases for each, are presented in Tables 4.1 and 4.2 below.

Bivariate Correlation

To answer the first research question, two exploratory factor analyses were conducted. Prior to conducting the factor analysis, a correlation matrix was produced in order to determine whether the variables were related. Tabachnick and Fidell (2007) state that if correlations do not exceed .30, factor analysis may not be an adequate statistical analysis. The correlation matrix in the current study yielded large bivariate correlations between many behavioral EF variables, supporting the use of factor analysis with this data. However, the correlations between the cognitive variables of EF were much lower, suggesting that factor analysis may not be the most appropriate methodology for these variables. Because the goal of the current study was to compare the factor structure of cognitive and behavioral measures of EF, it was determined to proceed with the factor analysis while noting concerns regarding weak correlation between cognitive factors as a limitation. Correlation data is outlined in Tables 4.3, 4.4, 4.5, and 4.6.

Exploratory Factor Analysis

Due to the large amount of missing data (23.42%), factor analysis was conducted using mPlus, which handles missing values within the analysis model, rather than replacing or imputing data (Collins, Shafer, & Kam, 2001). This method is termed full information maximum likelihood (FIML) and allows all available data to be included in the model. Although multiple imputation was initially proposed as a means of accounting for missing data, FIML was identified as more appropriate because it has been studied in the context of factor analysis whereas very little research has been conducted examining the use of data that has undergone multiple imputation in factor analysis. FIML and multiple imputation have been found to produce similar results.

Behavioral Factor Analysis

A factor analysis was conducted in order to examine the factor structure of behavioral measures of EF. mPlus was utilized to complete this phase of analysis due to its use of FIML. Twenty-two variables from behavioral measures of EF were factor analyzed using principle-axis factoring with oblique rotation. Oblique rotation was selected because correlations between the variables included in the analysis were generally above 0.3 (Meyers et al., 2006). Multiple phases of analysis were completed until a model in which all variables loaded on factors with factor loadings of .500 or higher was obtained. While conducting the analysis, factors with eigenvalues greater than 1.00 were retained. Variables with factor loadings less than .500 were considered criterion for removal from the next round of analyses, as well as variables that loaded on multiple factors. In cases where more than one variable met criteria for removal, the item with the lowest factor loadings across models was excluded during the next round. The following variables were removed from the analysis one-by-one and in this order due to failure to load on a factor and/or tendency to load on multiple factors: Daily Living, Atypicality, Adaptability, Withdrawal, and Inhibit. Therefore, only seventeen of the original twenty-two variables were included in the final model.

Table 4.1

Descriptive Statistics for Behavioral Variables

	Mean	SD	Skewness	Kurtosis	Missing
BASC DL	39.81	9.83	0.078	-0.860	140
BASC ADP	44.89	10.93	0.063	0.800	0
BASC AGG	51.71	10.16	1.120	-1.350	81
BASC ANX	54.76	13.00	0.463	1.050	0
BASC ATTN	60.11	8.98	-0.476	-0.245	16
BASC ATP	54.52	11.88	0.184	0.872	1
BASC CON	53.04	11.82	1.400	3.020	0
BASC DEP	55.14	11.97	0.982	0.673	0
BASC COM	42.56	10.38	0.096	-0.280	2
BASC HYP	56.68	10.85	0.462	-0.129	18
BASC LDR	43.81	9.00	0.368	-0.106	0
BASC SOC	46.93	10.55	0.103	-0.622	0
BASC SOM	52.55	13.43	1.140	1.130	1
BASC WTH	55.59	13.41	0.813	0.276	0
BR INH	58.21	12.68	0.337	0.478	8
BR SHT	59.01	12.99	0.400	-0.370	11
BR EC	57.00	12.95	0.410	-0.597	7
BR INT	59.51	10.90	-0.046	-0.697	6
BR WM	64.25	11.83	-0.224	-0.501	30
BR PL	62.24	11.38	0.055	-0.450	10
BR ORG	57.13	10.80	-0.045	-0.588	10
BR MON	60.33	10.56	-0.008	-0.428	19

Note. Scores are reported as *T*-scores. DL – Activities of Daily Living; ADP – Adaptability; AGG – Aggression; ANX – Anxiety; ATTN – Attention Problems; ATP – Atypicality; CON – Conduct Problems; DEP – Depression; COM – Communication; HYP – Hyperactivity; LDR – Leadership; SOC – Social Skills; SOM – Somatization; WTH – Withdrawal; INH – Inhibit; SHT – Shift; EC – Emotional Control; INT – Initiate; WM – Working Memory; PL – Plan/Organize; ORG – Organization of Materials; MON – Monitor

Table 4.2

Descriptive Statistics for Cognitive Variables

	Mean	SD	Skewness	Kurtosis	Missing
DK CW3	8.31	3.57	-0.409	-0.543	57
DK CW4	7.78	3.24	-0.369	-0.352	53
DK VL3	8.14	3.83	0.151	0.121	59
DK DF3	8.39	2.96	0.771	0.919	77
DK TMT4	7.17	3.66	-0.234	-1.000	55
DK TOW	9.38	3.17	0.314	0.934	116
DK WC	7.51	3.56	0.582	-0.140	100
NEPSY AS	8.60	4.28	0.202	-0.376	108
NEPSY AA	7.84	3.56	-0.338	-0.779	47
NEPSY RS	8.51	3.43	-0.267	-0.649	59
NEPSY INH	8.33	3.61	0.563	1.210	62
NEPSY WLI	8.36	3.61	0.237	-0.023	96
WJ CF	97.71	14.76	-0.274	0.830	85
WJ AS	95.88	15.85	-0.076	0.064	107
WJ NR	89.63	17.18	-0.163	-0.169	83

Note. Scores are reported as scaled scores (D-KEFS and NEPSY-2) and standard scores (WJ III COG NU). CW3 - Color Word Interference Condition 3; CW4 – Color Word Interference Condition 4; VL3 – Verbal Fluency Condition 3; DF3 – Design Fluency Condition 3; TMT 4 – Trail Making Test Condition 4; TOW – Tower; WC – Word Context; AS – Animal Sorting; AA – Auditory Attention; RS – Response Set; INH – Inhibition (Switching); WLI – World List Interference; CF – Concept Formation; AS – Analysis/Synthesis; NR – Numbers Reversed

Table 4.3

Pearson Product Moment Correlations Between BASC-2 Variables

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. AGG	1													
2. CON	.612**	1												
3. ANX	.109	.029	1											
4. DEP	.574**	.497**	.494**	1										
5. SOM	.153	.263**	.448**	.511**	1									
6. ATP	.440**	.367**	.365**	.525**	.234**	1								
7. WTH	.279**	.208**	.291**	.483**	.204**	.508**	1							
8. ADP	-.624**	-.383**	-.283**	-.557**	-.281**	-.446**	-.503**	1						
9. SOC	-.503**	-.395**	-.007	-.320**	-.047	-.314**	-.438**	.526**	1					
10. LDR	-.375**	-.259**	-.051	-.345**	.001	-.348**	-.461**	.485**	.631**	1				
11. DL	-.644**	-.513**	.074	-.486**	-.166	-.556**	-.195	.291	.463**	.380*	1			
12. COM	-.360**	-.290**	-.071	-.350**	-.098	-.512**	-.419**	.424**	.473**	.679**	.443**	1		
13. HYP	.464**	.389**	.334**	.420**	.258**	.380**	.260**	-.365**	-.297**	-.235**	-.485**	-.291**	1	
14. ATTN	.427**	.306**	.080	.236**	-.510	.398**	.266**	-.354**	-.371**	-.429**	-.612**	-.434**	.510**	1

Note. *$p < .05$ ** $p < .01$; AGG – Aggression; CON – Conduct Problems; ANX – Anxiety; DEP – Depression, SOM – Somatization; ATP – Atypicality; WTH – Withdrawal; ADP – Adaptability; SOC – Social Skills; LDR – Leadership; DL – Activities of Daily Living; COM – Communication; HYP – Hyperactivity; ATTN – Attention Problems.

Table 4.4

Pearson Product Moment Correlations Between BRIEF Variables

	1	2	3	4	5	6	7	8
1. INH	1							
2. SHT	.462**	1						
3. EC	.475**	.629**	1					
4. INT	.377**	.414**	.416**	1				
5. WM	.404**	.354**	.342**	.632**	1			
6. PL	.321**	.394**	.418**	.634**	.669**	1		
7. ORG	.298**	.234**	.235**	.456**	.464**	.627**	1	
8. MON	.503**	.432**	.427**	.543**	.572**	.666**	.400**	1

Note. **p* < .05 ** *p* < .01; INH – Inhibit; SHT – Shift; EC – Emotional Control; INT – Initiate; WM – Working Memory; PL – Planning/Organization; ORG – Organization of Materials; MON – Monitor

Table 4.5

Pearson Product Moment Correlations Between BASC-2 and BRIEF Variables

	BR INH	BR SHT	BR EC	BR INT	BR WM	BR PL	BR ORG	BR MON
BASC AGG	.472**	.393**	.397**	.235*	.298*	.335**	.394**	.236*
BASC CON	.450**	.158*	.237**	.283**	.247**	.314**	.306**	.309**
BASC ANX	.154*	.370**	.335**	.163*	.127	.017	.040	.065
BASC DEP	.388**	.414**	.564**	.365**	.246**	.308**	.289**	.321**
BASC SOM	.163*	.264**	.312**	.206**	.085	.176*	.187*	.095
BASC ATP	.363**	.437**	.381**	.350**	.333**	.183*	.102	.289**
BASC WTH	.249**	.461**	.348**	.384**	.170*	.237**	.184*	.227**
BASC ADP	-.439**	-.549**	-.480**	-.288**	-.297**	-.251**	-.301**	-.298**
BASC SOC	-.239**	-.247**	-.263**	-.285**	-.202*	-.211**	-.162*	-.210**
BASC LDR	-.242**	-.317**	-.225**	-.381**	-.340**	-.312**	-.213**	-.327**
BASC DL	-.387*	-.269	-.434*	-.665**	-.352	-.602**	-.188	-.360
BASC COM	-.259**	-.381**	-.214**	-.436**	-.364**	-.354**	-.244**	-.411**
BASC HYP	.546**	.351**	.400**	.361**	.263**	.243**	.294**	.368**
BASC ATTN	.341**	.236**	.187*	.441**	.501**	.331**	.274**	.339**

Note. Note. **p* < .05 ** *p* < .01; AGG – Aggression; CON – Conduct Problems; ANX – Anxiety; DEP – Depression; SOM – Somatization; ATP - Atypicality; WTH – Withdrawal; ADP – Adaptability; SOC – Social Skills; LDR – Leadership; DL – Daily Living; COM – Communication; HYP – Hyperactivity; ATTN – Attention; INH – Inhiit; SHT – Shift; EC – Emotional Control; INT – Initiate; WM – Working Memory; PL – Planning/Organization; ORG – Organization of Materials; MON – Monitor

Table 4.6

Pearson Product Moment Correlations Between Cognitive Variables

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. DK CW3	1													
2. DK CW4	.554**	1												
3. DK VL3	.171	.150	1											
4. DK DF3	.344**	.447**	-.103	1										
5. DK TMT4	.253*	.237*	.198*	.211*	1									
6. DK TOW	.231	.010	.080	.326*	.261	1								
7. DK WC	.055	-.083	.301*	-.222	.243*	.086	1							
8. NEP AS	-.157	-.242	.072	-.056	.079	.104	.170	1						
9. NEP AA	.126	.230*	-.087	.037	.066	.131	.104	.223	1					
10. NEP RS	.303**	.285*	-.113	.249*	.148	.207	.158	.020	.373*	1				
11. NEP INH	.291*	.205	-.002	.184	.395**	.349**	.116	.257	.102	.260*	1			
12. NEP WLI	.249	.244	.214	.323*	.271*	.392*	.277	.334	.073	.237	.413**	1		
13. WJ CF	.318*	.100	.249	.252	.450**	.185	.192	.453**	.191	.171	.376**	.510**	1	
14. WJ AS	.233	.027	.287	-.040	.608**	.054	.230	.654**	.125	.016	.345*	.217	.448**	1
15. WJ NR	.225	.208	.197	.142	.504**	.247	.312	.309	.191	.052	.448**	.420**	.311**	.482**

Note. *$p < .05$ ** $p < .01$; CW3 – Color Word Interference Condition 3; CW4 – Color Word Interference Condition 4; DF3 – Design Fluency; TMT4 – Trail Making Test Condition 4; TOW – Tower; WC – Word Context; AS – Animal Sorting; RS – Response Set; INH – Inhibition (Switching); WLI – Word List Interference; CF – Concept Formation; AS – Analysis/Synthesis; NR – Numbers Reversed

Table 4.7

Pearson Product Moment Correlations Between Cognitive Variables and BASC-2 Variables

	AGG	CON	ANX	DEP	SOM	ATP	WTH	ADP	SOC	LDR	DL	COM	HYP	ATTN
DK CW3	.076	.124	.092	.370	.064	.045	-.030	-.099	-.106	.191*	-.192	.020	.082	.136
DK CW4	-.005	-.036	.013	-.120	.015	-.102	-.110	.032	.012	.069	-.280	.147	-.132	-.039
DK VL3	-.105	.076	.096	.030	.125	-.055	.049	.031	-.020	.224*	.022	.122	.072	-.097
DK DF3	-.218	-.131	.018	-.109	-.005	-.264**	-.187	.052	.087	.082	.169	.173	-.167	-.101
DK TMT4	-.117	-.076	-.002	-.036	-.013	-.278**	-.054	-.022	-.036	-.028	-.058	.220*	.027	-.002
DK TOW	-.144	.087	-.122	-.066	-.076	-.193	-.225	.046	-.121	-.003	-.142	.146	-.127	.139
DK WC	.187	.136	.086	.106	.500	.008	.110	-.066	-.111	.117	.026	.181	.237*	.159
NEP AS	-.201	.009	-.044	-.058	.220	-.128	-.054	-.107	-.074	.190	.387	.069	.053	-.014
NEP AA	.002	.103	.014	-.149	-.125	-.145	-.085	.104	-.021	.154	-.078	.149	-.013	-.044
NEP RS	-.204	-.133	.158	-.115	-.208*	-.086	-.023	.031	.082	.118	.214	.126	-.120	-.129
NEP INH	-.076	-.150	.000	-.139	-.066	-.148	-.126	-.026	.106	.085	-.139	.121	.009	-.050
NEP WLI	-.103	.004	.032	-.206	.091	-.175	-.139	.028	-.032	.182	-.125	.305**	.097	.000
WJ CF	-.157	-.057	.046	.030	.036	.004	.041	-.092	.098	.200	.620	.206	.105	.034
WJ AS	-.025	-.098	.042	-.032	.068	-.274*	-.012	-.044	.047	.136	-.125	.153	.029	.075
WJ NR	-.126	.079	.231*	.155	.183	-.224*	.072	-.099	-.021	.112	-.171	.332**	.106	-.122

Note. *$p < .05$ ** $p < .01$; CW3 – Color Word Interference Condition 3; CW4 – Color Word Interference Condition 4; DF3 – Design Fluency; TMT4 – Trail Making Test Condition 4; TOW – Tower; WC – Word Context; AS – Animal Sorting; RS – Response Set; INH – Inhibition (Switching); WLI – Word List Interference; CF – Concept Formation; AS – Analysis/Synthesis; NR – Numbers Reversed; AGG – Aggression; CON – Conduct Problems; ANX – Anxiety; DEP – Depression; SOM – Somatization, ATP - Atypicality; WTH – Withdrawal; ADP – Adaptability; SOC – Social Skills; LDR – Leadership; DL – Daily Living; COM – Communication; HYP – Hyperactivity; ATTN – Attention

Table 4.8

Pearson Product Moment Correlations Between Cognitive Variables and BRIEF Variables

	INH	SHT	EC	INT	WM	PL	ORG	MON
DK CW3	.560	-.036	-.153	-.018	-.144	-.082	.260	.035
DK CW4	-.172	-.130	-.266**	.030	-.104	-.112	-.022	-.196*
DK VL3	.230	-.069	-.041	.084	-.106	.089	.150	.670
DK DF3	-.283**	-.181	-.273**	-.153	.025	-.195	-.106	-.134
DK TMT4	-.034	-.102	-.110	-.006	.002	.091	.124	.024
DK TOW	.210	-.165	-.022	.340	.150	.150	.194	.087
DK WC	.259*	-.075	.207	.199	.380	.161	.191	.071
NEP AS	-.009	-.202	.077	.640	-.070	.054	-.114	-.103
NEP AA	.007	-.118	-.091	-.159	-.080	-.088	.015	-.246**
NEP RS	-.055	-.042	-.118	-.200*	-.178	-.252**	-.102	-.190*
NEP INH	-.010	-.114	-.590	-.097	-.003	.017	-.052	.930
NEP WLI	-.110	-.158	-.118	-.203	-.111	-.122	-.152	.000
WJ CF	.039	-.060	.012	.028	-.077	.610	.700	.029
WJ AS	-.054	-.052	-.042	.012	-.005	.950	.370	-.100
WJ NR	-.011	-.071	-.046	-.084	-.058	-.010	.490	-.079

Note. $*p < .05$ $**p < .01$; CW3 – Color Word Interference Condition 3; CW4 – Color Word Interference Condition 4; DF3 – Design Fluency; TMT4 – Trail Making Test Condition 4; TOW – Tower; WC – Word Context; AS – Animal Sorting; RS – Response Set; INH – Inhibition (Switching); WLI – Word List Interference; CF – Concept Formation; AS – Analysis/Synthesis; NR – Numbers Reversed; INH – Inhiit; SHT – Shift; EC – Emotional Control; INT – Initiate; WM – Working Memory; PL – Planning/Organization; ORG – Organization of Materials; MON – Monitor

The first factor analysis yielded five factors explaining a total of 10.32% of the variance for the entire set of variables. Factor 1 was labeled Externalizing due to the high loadings of Aggression and Conduct Problems. The first factor explained 1.31% of the variance. Factor 2 was labeled Internalizing/Self-Regulation due to the high loadings of Anxiety, Depression, Somatization, Shift, and Emotional Control. The second factor explained 2.68% of the variance. Factor 3 was labeled Adaptive due to the high loadings of Communication, Social Skills, and Leadership. The third factor explained 1.97% of the variance. Factor 4 was labeled ADHD due to the high loadings of Hyperactivity and Attention. The fourth factor explained 1.27% of the variance. Factor 5 was labeled Metacognition due to the high loadings of Initiate, Working Memory, Planning, Organization of Materials, and Monitor. The fifth factor explained 3.09% of the variance.

Cognitive Factor Analysis

Prior to conducting a factor analysis on the cognitive measures of EF, variables were transformed to *Z*-scores using SPSS in order to convert them to the same measurement scale. This was completed due to the significant difference in range between scale and standard scores. Although some researchers suggest that the difference in scale may not affect results because factor analysis is based upon correlations, others argue that error variance differences between scales may result in bias (Reise, Waller, & Comrey, 2000).

Table 4.9

Factor Loadings of Behavioral Variables

	Externalizing	Self-Regulation	Adaptive	Metacognition	ADHD
BASC AGG	.649				
BASC CON	.814				
BASC ANX		.764			
BASC DEP		.586			
BASC SOM		.520			
BASC SOC			.598		
BASC LDR			.960		
BASC COM			.617		
BASC HYP					.510
BASC ATTN					.693
BR SHT		.508			
BR EC		.539			
BR INT				.573	
BR WM				.686	
BR PL				.972	
BR ORG				.604	
BR MON				.673	

Note. AGG – Aggression; CON – Conduct Problems; ANX – Anxiety; DEP – Depression; SOM – Somatization; SOC – Social Skills; LDR – Leadership; COM – Communication; HYP – Hyperactivity; ATTN – Attention Problems; SHT – Shift; EC – Emotional Control; INT – Initiate; WM – Working Memory; PL – Plan/Organize; ORG – Organization of Materials; MON – Monitor

A second factor analysis was conducted in order to examine the factor structure of cognitive measures of EF. mPlus was utilized to complete this phase of analysis due to its use of FIML. Fifteen variables from the cognitive measures of EF were factor analyzed using principle-axis factoring with oblique rotation. Although cognitive measures of EF had somewhat weak correlations, oblique rotation was utilized to increase comparability between the behavioral and cognitive factor analyses. Multiple phases of analysis were completed until a model in which all variables loaded on factors with factor loadings of .500 or higher was obtained. While conducting the analysis, factors with eigenvalues greater than 1.00 were retained. Variables with factor loadings less than .500 were considered criterion for removal from the next round of analyses, as well as variables that loaded on multiple factors. In cases where more than one variable met criteria for removal, the item with the lowest factor loadings across models was excluded during the next round. The following variables were removed from the analysis one-by-one and in this order due to failure to load on a factor and/or tendency to load on multiple factors: Tower, Verbal Fluency 3, Auditory Attention, Inhibition (Switching), Response Set, Word List Interference, and Animal Sorting. Therefore, only eight of the original fifteen variables were included in the final factor solution.

The analysis yielded only two factors explaining a total of 3.46% of the variance of the entire set of variables. Factor 1 was labeled Shifting due to the high loadings of Color Word Interference Condition 3, Color Word Interference Condition 4, and Design Fluency Condition 3. The first factor explained 1.42% of the variance. Factor 2 was labeled Reasoning due to the high loadings of Trail Making Test Condition 4, Word Context, Concept Formation, Analysis/Synthesis, and Numbers Reversed. The second factor explained 2.04% of the variance. Overall, this model was not statistically significant at the .05 level; therefore, reliability statistics were completed in order to investigate the strength of the model.

Table 4.10

Factor Loadings of Cognitive Variables

	Shifting	Reasoning
DK CW3	.558	
DK CW4	.862	
DK DF3	.500	
DK TMT4		.715
DK WC		.507
WJ CF		.536
WJ AS		.798
WJ NR		.599

Note. CW3 – Color Word Interference Condition 3; CW4 – Color Word Interference Condition 4; DF3 – Design Fluency Condition 3; TMT4 – Trail Making Test Condition 4; WC – Word Context; CF – Concept Formation; AS – Analysis/Synthesis; NR – Numbers Reversed

Reliability Analysis

An important final step in EFA is investigating the reliability of each of the factors. The seventeen variables making up the five behavioral factors and the eight variables making up the two cognitive factors were subjected to a reliability analysis to determine the reliability of each factor. Cronbach's alpha is used as a reliability estimate (Field, 2009; Meyers et al., 2006). A reliability coefficient of .70 or higher is considered acceptable in most social science studies.

In order to complete this part of the analysis, mean scores for each factor were first calculated using SPSS. A mean score for each participant for each factor was calculated by averaging the scores across the variables making up each factor. In order to account for missing data, parameters were set to ensure that means were only calculated if the majority of scores for each factor were available for the participants. For example, when calculating the mean of the Internalizing factor, which included five variables, three of the five variables had to have complete data in order for a mean to be calculated. For factors made up of only two variables, both variables needed complete data in order for the mean to be calculated. Because of this parameter, some cases did not have mean values calculated and thus had missing data. However, this was not anticipated to affect results of reliability analyses because the data was MCAR and there was an adequate number of cases with data.

Behavioral factors. Results revealed that the two variables making up the Externalizing factor demonstrated high reliability (Cronbach's alpha = .751). The five variables making up the Internalizing/Self-Regulation factor demonstrated high reliability (Cronbach's alpha = .792). The three variables making up the Adaptive factor demonstrated high reliability (Cronbach's alpha = .809). The two variables making up the ADHD factor demonstrated moderate reliability (Cronbach's alpha = .668). The five variables making up the Metacognition factor demonstrated high reliability (Cronbach's alpha = .868). Overall, the factors obtained during the factor analysis demonstrated adequate reliability with the exception of the ADHD factor, which fell below the cutoff.

Cognitive factors. Results revealed that the three variables making up the Shifting factor demonstrated high reliability (Cronbach's alpha = .739). The five variables making up the Reasoning factor demonstrated moderate reliability (Cronbach's alpha = .699). Overall, the factors obtained during the factor analysis demonstrated adequate reliability although it should be noted that the reliability coefficient for the Reasoning factor fell just below the cutoff.

Correlation Between Factors

The third research question was addressed by creating a correlation matrix of the mean scores for each factor, which were computed during the previous stage of analysis. Pearson product-moment correlations were used to determine the relationships between behavioral and cognitive factors of EF as obtained during exploratory factor analysis. There were no statistically significant correlations observed between any of the behavioral and cognitive EF factors. Most of the behavioral EF factors were strongly correlated with one another, while none of the cognitive EF factors were correlated with one another. Please refer to table 4.11 for more information.

Table 4.11

Correlations Between Behavioral and Cognitive EF Factors

	1	2	3	4	5	6	7
1. Adaptive	1						
2. ADHD	-.474**	1					
3. Internalizing	.529**	-.485**	1				
4. Externalizing	.446**	-.298**	.412**	1			
5. Metacognition	.432**	-.442**	.489**	.422**	1		
6. Reasoning	-.252	.182	.027	.049	.081	1	
7. Shifting	-.075	.109	-.020	-.111	-.120	.221	1

Note. *$p < .05$ **$p < .01$

Linear Regression

In order to investigate the third research question, a series of linear regression analyses were carried out. Initially, it was proposed that the variables comprising each behavioral factor would be used to predict the variables comprising cognitive factors that correlated strongly with the behavioral factor. However, no significant correlations were obtained for the behavioral and cognitive factors, meaning they were not related. Instead, a series of linear regression analyses were conducted to determine whether the behavioral variables making up each obtained factor predicted any of the cognitive variables. This was completed with the goal of identifying aspects of EF as measured by behavioral rating scales that predict performance on cognitive measures of EF.

Five regression analyses (using each of the five behavioral factors) were completed for each cognitive variable, resulting in a total of 75 regression analyses. This large number of analyses was completed to ensure that a majority of cases were included and to reduce the number of variables used to predict the outcome variable. These analyses were completed using Stata in order to obtain information regarding multicollinearity. Furthermore, analyses were completed using incomplete data and cases were excluded listwise, meaning if they did not contain data for all variables included in the analysis, they were not included in the analysis. Due to the large number of regression analyses completed, only significant results will be discussed below. As anticipated based on the results of the correlation between factor means, most behavioral variables did not predict cognitive variables.

D-KEFS

Color Word Interference 3. Results indicated that Emotional Control ($\beta = -.077$, $p = .030$) explained 5.2% of the variance in Color Word Interference 3, $R^2 = .052$, F (5, 109) = 1.18, $p = .321$. While more clinically significant ratings of Emotional Control predicted lower scores on Color Word Interference 3, the overall model for this factor was not significant. Social Skills (β = -.106, p = .005) and Leadership (β = .195, p = .000) explained 12.2% of the variance in Color Word Interference 3, R^2 = .112, F (3, 113) = 5.24, p = .002. Higher ratings of Social Skills predicted lower scores on Color Word Interference 3, while higher ratings of leadership predicted higher scores on Color Word Interference 3. The Adaptive factor appears to predict performance on this test.

Table 4.12

Significant Regression Analyses for Color Word Interference 3

Factor	Predictors	R^2	p	β	SE	p
Internalizing		.052	.321			
	BASC ANX			.020	.031	.536
	BASC DEP			.025	.037	.513
	BASC SOM			.012	.033	.724
	BR SHT			.022	.035	.534
	BR EC			-.077	.035	.030
Adaptive		.122	.002			
	BASC COM			.047	.040	.245
	BASC SOC			-.106	.037	.005
	BASC LDR			.196	.054	.000

Note. ANX – Anxiety; DEP – Depression; SOM – Somatization; SHT – Shift; EC – Emotional Control; COM – Communication; SOC – Social Skills; LDR – Leadership

Color Word Interference 4. Emotional Control (β = -.086, p = .006) explained 10.3% of the variance in Color Word Interference 4, R^2 = .103, F (5, 108) = 2.48, p = .036. More clinically significant ratings of Emotional Control predicted lower scores on Color Word Interference 4. Initiate (β = .087, p = .036) explained 8.9% of the variance in Color Word Interference 4, R^2 = .089, F (5, 82) = 1.61, p = .168. While more clinically significant ratings of initiate predicted higher scores on Color Word Interference 4, the overall model for this factor was not significant. The Internalizing factor appears to predict performance on this test.

Table 4.13

Significant Regression Analyses for Color Word Interference 4

Factor	Predictors	R^2	*p*	β	SE	*p*
Internalizing		.103	.036			
	BASC ANX			.019	.029	.507
	BASC DEP			-.017	.033	.614
	BASC SOM			.032	.030	.285
	BASC SHT			.017	.032	.603
	BASC EC			-.086	.031	.006
Metacognition		.089	.168			
	BR INT			.087	.041	.036
	BR WM			-.467	.041	.257
	BR PL			.002	.049	.970
	BR ORG			-.008	.041	.855
	BR MON			-.072	.045	.116

Note. ANX – Anxiety; DEP – Depression; SOM – Somatization; SHT – Shift; EC – Emotional Control; INT – Initiate; WM – Working Memory; PL – Plan/Organize; ORG – Organization of Materials; MON – Monitor

Verbal Fluency 3. Social Skills (β = -.073, $p = .047$) and Leadership (β = .154, $p = .005$) explained 8.7% of the variance in Verbal Fluency 3, $R^2 = .087$, F (3, 111) = 3.51, $p = .018$. Higher ratings of Social Skills predicted lower scores on Verbal Fluency while higher ratings of Leadership predicted higher scores on Verbal Fluency. The Adaptive factor appears to predict performance on this test.

Table 4.14

Significant Regression Analyses for Verbal Fluency 3

Factor	Predictors	R^2	*p*	β	SE	*p*
Adaptive		.087	.018			
	BASC COM			-.016	.039	.685
	BASC SOC			-.074	.037	.047
	BASC LDR			.154	.054	.005

Note. COM – Communication; SOC – Social Skills; LDR – Leadership

Design Fluency 3. Emotional Control (β = -.069, $p = .039$) explained 9.5% of the variance in Design Fluency 3, $R^2 = .095$, F (5, 88) = 1.85, $p = .112$. While more clinically significant ratings of Emotional Control predicted lower scores on Design Fluency 3, the overall model for this factor was not significant. No behavioral factors predicted performance on this task.

Table 4.15

Significant Regression Analyses for Design Fluency 3

Factor	Predictors	R^2	*p*	β	SE	*p*
Internalizing		.095	.112			
	BASC ANX			.018	.035	.617
	BASC DEP			-.002	.036	.952
	BASC SOM			.019	.033	.569
	BR SHT			-.010	.035	.784
	BR EC			-.069	.033	.039

Note. ANX – Anxiety; DEP – Depression; SOM – Somatization; SHT – Shift; EC – Emotional Control

Trail Making Test 4. Communication (β = .149, $p = .000$) and Leadership (β = -.112, $p = .048$) explained 10.4% of the variance in Trail Making Test 4, $R^2 = .103$, F (3, 115) = 4.43, $p = .006$. Higher ratings of communication predicted higher scores on Trail

Making Test 4, while higher ratings of Leadership predicted lower scores on Trail Making Test 4. The Adaptive factor appears to predict performance on this test.

Table 4.16

Significant Regression Analyses for Trail Making Test 4

Factor	Predictors	R^2	*p*	β	SE	*p*
Adaptive		.104	.006			
	BASC COM			.149	.041	.000
	BASC SOC			-.028	.040	.476
	BASC LDR			-.113	.057	.048

Note. COM – Communication; SOC – Social Skills; LDR - Leadership

Tower. Conduct Problems (β = .119, p = .038) explained 19.1% of the variance in Tower, R^2 = .191, F (2, 23) = 2.72, p = .087. While more clinically significant ratings of Conduct Problems predicted higher scores on Tower, the overall model for this factor was not significant. Additionally, the number of cases containing the variables comprising this model was very low (n = 26); thus, results of this analysis should be interpreted with caution. No behavioral factors predicted performance on this test.

Table 4.17

Significant Regression Analyses for Tower

Factor	Predictors	R^2	*p*	β	SE	*p*
Externalizing		.191	.087			
	BASC AGG			-.138	.068	.056
	BASC CON			.119	.054	.038

Note. AGG – Aggression; CON – Conduct Problems

Word Context. Shift (β = -.103, p = .020) and Emotional Control (β = .112, p = .015) explained 12.7% of the variance in Word Context, R^2 = .127, F (5, 67) = 1.95, p =

.098. While more clinically significant ratings of Shift predicted lower scores on Word Context, more clinically significant ratings of Emotional Control predicted higher scores on Word Context. Additionally, the overall model for this factor was not significant. Initiate (β = .117, *p* = .034) explained 15.6% of the variance in Word Context, R^2 = .156, *F* (5, 45) = 1.66, *p* = .163. While more clinically significant ratings of Initiate predicted higher scores on Word Context, the overall model for this factor was not significant. No behavioral factors predicted performance on this test.

Table 4.18

Significant Regression Analyses for Word Context

Factor	Predictors	R^2	*p*	β	SE	*p*
Internalizing		.127	.098			
	BASC ANX			.035	.043	.412
	BASC DEP			.009	.042	.832
	BASC SOM			-.034	.041	.417
	BR SHT			-.103	.043	.020
	BR EC			.113	.045	.015

Note. ANX – Anxiety; DEP – Depression; SOM – Somatization; SHT – Shift; EC – Emotional Control

NEPSY-2

Animal Sorting. Somatization (β = .113, *p* = .044) explained 18.0% of the variance in Animal Sorting, R^2 = .180, *F* (5, 59) = 2.59, *p* = .035. More clinically significant scores of Somatization predicted higher scores on Sorting. Social Skills (β = -.120, *p* = .048) and Leadership (β = .235, *p* = .018) explained 9.9% of the variance in Animal Sorting, R^2 = .099, *F* (3, 64) = 2.35, *p* = .081. While, higher ratings of Social Skills predicted lower scores on Animal Sorting and higher ratings of Leadership predicted higher scores on Animal Sorting, the overall model for this factor was not significant. Planning (β = .207, p = .037), Organization of Materials (β = -.183, p = .014), and Monitor (β = -.200, p = .019) explained 22.2% of the variance in Animal Sorting, R^2 = .222, F (5, 46) = 2.63, p = .036. More clinically significant ratings of planning predicted higher scores on Animal Sorting, while more clinically significant scores on Organization of Materials and Monitor predicted lower scores on Animal Sorting. The Internalizing and Metacognition factors appears to predict performance on this test.

Table 4.19

Significant Regression Analyses for Animal Sorting

Factor	Predictors	R^2	p	β	SE	p
Internalizing		.018	.035			
	BASC ANX			-.022	.049	.647
	BASC DEP			-.086	.068	.209
	BASC SOM			.113	.055	.044
	BR SHT			-.098	.051	.060
	BR EC			.091	.055	.105
Adaptive		.099	.081			
	BASC COM			-.059	.072	.413
	BASC SOC			-.120	.060	.048
	BASC LDR			.235	.097	.018
Metacognition		.222	.036			
	BR INT			.126	.076	.106
	BR WM			-.108	.079	.176
	BR PL			.206	.096	.037
	BR ORG			-.183	.072	.014
	BR MON			-.200	.082	.019

Note. ANX – Anxiety; DEP – Depression; SOM – Somatization; SHT – Shift; EC – Emotional Control; COM – Communication; SOC – Social Skills; LDR – Leadership; INT – Initiate; WM – Working Memory; PL – Plan/Organize; ORG – Organization of Materials; MON – Monitor

Auditory Attention. Monitor (β = -.105, *p* = .031) explained 7.2% of the variance in Auditory Attention, R^2 = .072, *F* (5, 90) = 1.40, *p* = .230. More clinically significant ratings on Monitor predicted lower scores on Auditory Attention. However, the overall model for this factor was not significant. No behavioral factors predicted performance on this test.

Table 4.20

Significant Regression Analyses for Auditory Attention

Factor	Predictors	R^2	*p*	β	SE	*p*
Metacognition		.072	.231			
	BR INT			.053	.047	.257
	BR WM			-.042	.047	.372
	BR PL			.061	.062	.330
	BR ORG			.023	.045	.607
	BR MON			-.105	.048	.031

Note. INT – Initiate; WM – Working Memory; PL – Plan/Organize; ORG – Organization of Materials; MON – Monitor

Response Set. Anxiety (β = .106, *p* = .001) and Somatization (β = -.063, *p* = .022) explained 14.5% of the variance in Response Set, R^2 = .145, *F* (5, 104) = 3.52, *p* = .006. More clinically significant ratings of anxiety predicted higher scores on Response Set, while more clinically significant scores of Somatization predicted lower scores on Response Set. The Internalizing factor appears to predict performance on this test.

Word List Interference. Depression (β = -.104, *p* = .043) explained 10.9% of the variance in Word List Interference R^2 = .109, *F* (5, 67) = 1.64, *p* = .162. While more clinically significant ratings of Depression predicted lower scores on Word List Interference, the overall model for this factor was not significant. Communication (β =

.116, p = .029) explained 12.7% of the variance in Word List Interference, R^2 = .127, F (3, 74) = 3.60, p = .017. Higher ratings of communication predicted higher scores on Word List Interference. The Adaptive factor appears to predict performance on this test.

Table 4.21

Significant Regression Analyses for Response Set

Factor	Predictors	R^2	p	β	SE	p
Internalizing		.145	.006			
	BASC ANX			.106	.030	.001
	BASC DEP			-.046	.043	.286
	BASC SOM			-.063	.027	.022
	BR SHT			-.007	.033	.830
	BR EC			-.018	.034	.584

Note. ANX – Anxiety; DEP – Depression; SOM – Somatization; SHT – Shift; EC – Emotional Control

Table 4.22

Significant Regression Analyses for Word List Interference

Factor	Predictors	R^2	p	β	SE	p
Internalizing		.109	.162			
	BASC ANX			.036	.035	.306
	BASC DEP			-.104	.050	.043
	BASC SOM			.062	.039	.117
	BR SHT			-.033	.047	.479
	BR EC			.017	.046	.714
Adaptive		.127	.017			
	BASC COM			.116	.052	.029
	BASC SOC			-.082	.049	.096
	BASC LDR			.042	.066	.528

Note. ANX – Anxiety; DEP – Depression; SOM – Somatization; SHT – Shift; EC – Emotional Control; COM – Communication; SOC – Social Skills; LDR – Leadership

Inhibition (Switching). No behavioral measures or factors were found to significantly predict performance on Inhibition (Switching), suggesting that behavioral measures of EF are a poor predictor of performance on this test.

WJ III COG NU

Communication (β = .823, p = .001) explained 15.5% of the variance in Numbers Reversed, R^2 = .155, F (3, 88) = 5.37, p = .002. Higher ratings of communication predicted higher scores on Numbers Reversed. The Adaptive factor appears to predict performance on this test. No behavioral measures or factors were found to significantly predict performance on Concept Formation or Analysis/Synthesis.

Table 4.23

Significant Regression Analyses for Numbers Reversed

Factor	Predictors	R^2	p	β	SE	p
Adaptive		.155	.002			
	BASC COM			.823	.229	.001
	BASC SOC			-.257	.215	.236
	BASC LDR			-.222	.296	.455

Note. COM – Communication; SOC – Social Skills; LDR – Leadership

Summary

This chapter attempted to explore and compare the factor structures of cognitive and behavioral measures of EF. Prior to completing primary statistical analyses, a missing data analysis was completed. To reduce the amount of missingness, variables were excluded if they were missing data for over two-thirds of the participants. Cases were excluded if they did not have data for behavioral variables or were missing two-thirds of the cognitive variables. Using these criteria, the final dataset included 22 behavioral variables and 15 cognitive variables with 176 participants between the ages of 8 and 16 with a variety of clinical diagnoses. The majority of the participants were Caucasian males and the average IQ score for the sample was 91.80. The final dataset with 176 participants had a total missing data percentage of 23.42%. Using Little's MCAR test (Little, 1988), the data was found to be missing completely at random (MCAR). Therefore, full information maximum likelihood and pairwise deletion were identified as appropriate means of addressing the missing data. The data was also analyzed to ensure that all assumptions were met. Three extreme outliers were detected but were included in final analyses because they were not found to significantly impact results. Most variables demonstrated normality but certain behavioral variables exhibited slight skewness to the right; however, the overall distribution was considered adequate for completing the analyses. Variance inflation factors (VIF) were calculated to ensure that multicollinearity did not bias regression analyses. All VIF scores were within the normal range, suggesting multicollinearity was not an issue.

Pearson product moment correlation analyses were completed to investigate the relationship between all variables. Moderate to strong correlations were observed between many of the behavioral variables, while weak correlations were observed between cognitive variables. Correlation values were extremely low between most behavioral and cognitive variables.

Two exploratory factor analyses were completed using full information maximum likelihood to account for missing data. Multiple phases of each analysis were completed until models in which all variables had factor loadings of .500 or higher were obtained.

Specific guidelines were followed for eliminating variables from each analysis one at a time. The variables included in the cognitive analysis were transformed to Z-scores so they were all on the same scale. A five factor model was obtained during the behavioral EFA, while a two factor model was obtained during the cognitive EFA. Behavioral factors were named Externalizing, Internalizing/Self-Regulation, Adaptive, Metacognition, and ADHD. Cognitive factors were name Shifting and Reasoning. Reliability analyses were computed to investigate the reliability of each factor. Most factors exhibited Cronbach's alpha scores of .70 or higher with the exceptions of ADHD and Reasoning, which were slightly below the .70 cutoff.

Pearson product moment correlations were then computed to determine the relationships between behavioral and cognitive factors of EF obtained during the factor analysis phase. Mean scores for each factor were created for each participant. If participants did not have a majority of the scores making up each factor, a mean score was not calculated for that factor. Correlations between mean scores revealed moderate ($r > .4$) correlations between most behavioral variables and weak correlations ($r < .2$) between most cognitive variables and between cognitive and behavioral variables.

Finally, five regression analyses (i.e., one for each of the five behavioral factors) were completed for each cognitive variable to determine whether the behavioral variables making up each factor predicted the cognitive variables. The Adaptive and Internalizing factors appeared to best predict performance on the D-KEFS variables, while the Adaptive, Internalizing, and Metacognition factors best predicted performance on the

NEPSY-2 variables. Overall, the factors were poor predictors of performance on the WJ III COG NU.

CHAPTER V

DISCUSSION

The current study investigated the factor structures of behavioral and cognitive measures of EF in a mixed clinical sample using the BASC-2, BRIEF, D-KEFS, NEPSY-2, and WJ III COG NU. Two exploratory factor analyses were completed in an attempt to identify the domains of EF that are measured by both behavioral (i.e., rating-based) and cognitive (i.e., performance-based) measures of EF. The obtained factors were then compared using correlation and regression analyses in order to determine whether behavioral measures of EF predict performance on cognitive measures of EF.

Purpose of the Study

The primary purpose of this study was to determine the relationship between cognitive and behavioral measures of EF in a mixed clinical sample by exploring the following three research questions:

1. What factor structures are obtained for behavioral measures of EF (i.e., parent ratings) and cognitive measures of EF (i.e., performance-based)?
2. Are the factor structures obtained for behavioral EF and cognitive EF comparable? In other words, do the behavioral EF factors correlate with the cognitive EF factors?
3. Finally, do the subtests loading on behavioral EF factors predict those loading on correlated cognitive EF factors?

The Integrated SNP/CHC model (Miller, 2013) was used as the basis for hypotheses. This model suggests that the construct of EF is composed of four domains, which are cognitive flexibility, concept recognition and generation, problem-solving, fluid reasoning and planning, and response inhibition. Other researchers (Barkley, 2012b; McCloskey & Perkins, 2012) suggest that attention and working memory are other key aspects of EF. Miller (2013) classifies these abilities as cognitive facilitators/inhibitors; however, because certain variables selected for inclusion in this study are thought to measure attention and working memory, it was hypothesized that factor analyses using cognitive measures of EF would reveal six factors made up of the four domains of EF as well as two domains of cognitive facilitators/inhibitors. It was further hypothesized that behavioral measures of EF would reveal two factors based on previous factor analytic studies using the BRIEF, which is divided into the domains of Self-Regulation and Metacognition (Gioia et al., 2000).

In regards to the second research question, it was hypothesized that correlations between some of the factor loadings obtained for cognitive and behavioral measures of EF would be correlated. Studies have shown that ratings on the BRIEF tend to correlate with ratings on the BASC and with classical measures of EF (Reynolds & Kamphaus, 2004). Furthermore, correlations have been observed between performance on various NEPSY-2 tasks and ADHD symptomatology as rated on the Brown ADD Scales (Brown, 2001). In other words, as ADHD symptomatology decreased, performance on the NEPSY-2 improved. In contrast, other researchers have found very weak correlations between parent ratings of EF and scores on EF tasks (Wilson, 1998). Differences in findings likely result from the complexity of the construct of EF as well as differences in types of EF measures. Although research is unclear in this area, findings suggest that some degree of a relationship exists between behavioral and cognitive measures of EF.

Lastly, it was hypothesized that the relationship between behavioral and cognitive measures of EF would be strong enough to allow behavioral measures of EF to predict performance on related cognitive measures of EF. This hypothesis was based on the idea that EF strongly influences behaviors, meaning behaviors serve as physical manifestations of inner functioning. This phenomenon was described by Barkley (2012b), who explained that children with ADHD and autism spectrum disorder exhibit specific patterns of behaviors that often reveal underlying dysfunction of EF systems. Because of this, it was anticipated that clinically significant ratings of behaviors related to EF dysfunction, as rated by parents and caregivers, would predict performance on similar EF scales. For example, a child with clinically significant ratings of Inhibition on the BRIEF or Hyperactivity on the BASC-2 might exhibit poor performance on Color Word Interference 4, which is purported to measure Inhibition.

Summary of Results

After reducing the sample size to account for missingness, completing a missing data analysis, and ensuring assumptions were met, Pearson product moment correlation analyses were completed to investigate the relationship between all behavioral and cognitive variables. In general, BASC-2 variables were strongly correlated with one another, with the strongest correlations ($r > .6$) observed between Aggression and Conduct Problems, Aggression and Adaptability (negative correlation), Aggression and

Daily Living (negative correlation), Social Skills and Leadership, Communication and Leadership, and Hyperactivity and Attention. When analyzing the BRIEF variables, strong correlations ($r > .6$) were observed between Emotional Control and Shift, Working Memory and Initiate, Working Memory and Plan/Organize, Organization of Materials and Plan/Organize, and Monitor and Plan/Organize. Additionally, a strong negative correlation was observed between Daily Living on the BASC-2 and Plan/Organize on the BRIEF. Other variables between the two behavioral rating scales demonstrated weak to moderate correlations. As anticipated, positive correlations were observed between the clinical scales on the BASC-2 and all scales on the BRIEF, while negative correlations were observed between the adaptive scales on the BASC-2 and all scales on the BRIEF. In other words, more significant impairments in all areas assessed by the BASC-2 tended to correlate with more significant impairments in all areas assessed by the BRIEF, while stronger adaptive functioning tended to relate to less significant impairment in all areas assessed on the BRIEF.

The cognitive variables as measured by the D-KEFS, NEPSY-2, and WJ III COG NU demonstrated weak to moderate correlations with one another. Strong correlations ($r > .6$) were observed between Analysis/Synthesis and Trail Making Test 4 and between Analysis/Synthesis and Animal Sorting. When examining correlations between behavioral and cognitive variables, most correlations were weak to very weak. However, moderate to strong ($r > .5$) correlations were observed between Word Context and Somatization and between Concept Formation and Activities of Daily Living. These results suggest that behavioral measures of EF tend to assess similar aspects of EF or assess an overall construct of EF, whereas cognitive measures of EF tend to assess different aspects of EF or effectively differentiate between specific cognitive abilities. Furthermore, behavioral and cognitive measures of EF appear divergent.

Exploratory Factor Analysis

The next phase of analysis involved the completion of two exploratory factor analyses using full information maximum likelihood (FIML) to account for missing data. Multiple phases of each analysis were completed until models in which all variables had factor loadings of .500 or higher were obtained. Variables were removed from the analysis one at a time following specific guidelines, including factor loadings less than .500 and loading on multiple factors. Daily Living, Atypicality, Adaptability, Withdrawal, and Inhibit were removed from the analysis of behavioral variables due to low factor loadings and dual loadings. Tower, Verbal Fluency 3, Auditory Attention, Inhibition (Switching), Response Set, Word List Interference, and Animal Sorting were removed from the analysis of cognitive variables due to low factor loadings and dual loadings. Therefore, only seventeen of the initial twenty-two behavioral variables were included in the final model for behavioral measures of EF and only eight of the initial fifteen variables were included in the final model for cognitive measures of EF.

It was initially anticipated that a two factor solution would emerge for behavioral variables and a six factor solution would emerge for cognitive variables. However, a five factor model was obtained for behavioral measures and a two factor model was obtained for cognitive measures. These models only explained 10.32% and 3.46% of the overall variance, respectively. In social sciences, solutions are generally only considered satisfactory when they explain at least 60% of the total variance (Hair, Black, Babin, & Anderson, 2010). However, lower percentages may be considered more interpretable when latent variables, such as those making up the construct of EF, are being investigated. Additionally, the model obtained during the behavioral factor analysis was statistically significant. Although the model obtained during the cognitive factor analysis did not reach significance, reliability statistics support the reliability of each factor obtained. In other words, the latent variables observed during this phase of analysis appear to have converged to some degree. Therefore, the results of these factor analyses have the potential to inform future research exploring the subcomponents of EF.

The first factor extracted during the factor analysis for behavioral measures of EF included Aggression and Conduct Problems. This factor was named Externalizing because these two scales make up the externalizing scale on the BASC-2. The second extracted factor included Anxiety, Depression, Somatization, Shift, and Emotional Control. This was the only behavioral factor that included variables from both the BASC-2 and BRIEF. It was named Internalizing/Self-Regulation because it included variables from the Internalizing scale on the BASC-2 and Self-Regulation scale on the BRIEF. The third factor included Social Skills, Leadership, and Communication and was named Adaptive because the variables included in this factor are part of the Adaptive scales on the BASC-2. The fourth factor included Initiate, Working Memory, Plan/Organize, Organization of Materials, and Monitor. This factor was named Metacognition because all of these variables make up the Metacognition scale on the BRIEF. Finally, the fifth factor included Hyperactivity and Attention Problems and was named ADHD because ADHD is defined by hyperactivity and deficits in attention.

Behavioral factors tended to load together in correspondence with previous factor analytic studies used to create the subscales of the BASC-2 and BRIEF (Reynolds & Kamphaus, 2004; Gioia et al., 2000). Results of this analysis support the divergence of these variables into separate scales. Additionally, there appears to be some convergence of the Internalizing and Self-Regulation scales of the BASC-2 and BRIEF. The second factor suggests that internalizing symptoms related to anxiety, depression, and somatization as measured by the BASC-2 tend to overlap with deficits in the EF capacities of shifting and emotional control as measured by the BRIEF. This is consistent with past research demonstrating that individuals with depression tend to exhibit poor emotional and behavioral regulation (Hosenbocus & Chahal, 2012) while individuals with anxiety characterized by obsessive thoughts and compulsive behaviors, tend to exhibit deficits in cognitive shifting and inhibition (Hunter & Sparrow, 2012).

The first factor extracted during the factor analysis of cognitive measures of EF included Color Word Interference Condition 3, Color Word Interference Condition 4, and Design Fluency Condition 3. This factor was named Shifting because all three of these tests measure the EF subcomponent of set-shifting as outlined by Miller (2013). The second extracted factor included Trail Making Test Condition 4, Word Context, Concept Formation, Analysis/Synthesis, and Numbers Reversed and was named Reasoning because each of these tests measures aspects of reasoning, including inductive and deductive reasoning, according to Miller (2013).

As mentioned previously, the first three variables of the D-KEFS likely loaded together because they assess the same subcomponent of EF (i.e., set-shifting). However, Trail Making Test Condition 4, which is also considered a measure of shifting, loaded on a separate factor. Given that the D-KEFS was designed to include various single tests of EF, rather than as a means of assessing EF holistically (Golden & Freshwater, 2002), it is not surprising that the subtests loaded on different factors. It is possible that Trail Making Test Condition 4 loaded on a separate factor because it is a less pure measure of shifting than the other three tests, requiring use of a complex system of abilities, including visual attention and working memory. In other words, this test, along with Numbers Reversed, which measures Working Memory, may have loaded on the reasoning factor because they require more complex processes than those loading on the shifting factor. In fact, reasoning and working memory skills tend to develop after shifting skills, which are considered less complex (Hunter & Sparrow, 2012). In this way, the factor analysis of cognitive measures of EF may have separated out EF variables by complexity.

Correlation Between Factors

Pearson product moment correlations were then computed using mean scores for each factor to determine the relationships between behavioral and cognitive factors of EF obtained during the factor analysis phase. Correlations between behavioral factors of EF were moderate ($r > .4$) with the exception of the Self-Regulation and Adaptive factors which were weakly correlated ($r = -.298$). As anticipated, the Adaptive factor was negatively correlated with other behavioral factors, which were positively correlated with one another, suggesting that higher ratings on the Adaptive factor were associated with less clinically significant ratings on the other behavioral factors. The correlation between the cognitive factors was weak (r = .221) and correlations between cognitive and behavioral factors were very weak (r < .2) with the exception of Reasoning and Externalizing, which were weakly correlated (r = -.254). This finding suggests that stronger performance on the Reasoning subtests is slightly associated with less clinically significant ratings on the Externalizing factor. The weak correlations observed between behavioral and cognitive factors stands in contrast to the hypothesis for the second research question, which anticipated that significant correlations would exist between some behavioral and cognitive factors.

Linear Regression

Lastly, a series of seventy-five linear regression analyses were completed to investigate the ability to predict performance on each cognitive measures of EF with each factor obtained during the behavioral factor analysis. In general, the factors were somewhat poor predictors of performance, which stands in contrast to the hypothesis for the third research question; however, a few subtests and factors predicted performance on certain cognitive measures. Specifically, Emotional Control, Social Skills, and Leadership predicted performance on Color Word Interference 3. Impairments in Emotional Control appear to predict lower scores on this subtest, while higher adaptive functioning in the areas of Social Skills and Leadership predict higher scores. Emotional Control and Initiate predicted performance on Color Word Interference 4. More clinically significant ratings of Emotional Control and Initiate predicted higher scores on this task. On the other hand, more clinically significant ratings of Emotional Control predicted lower scores on Design Fluency 3. Social Skills and Leadership predicted performance on Verbal Fluency 3 with higher ratings of both predicting higher scores on Verbal Fluency 3. Communication and Leadership predicted performance on Trail Making Test 4; however, higher ratings of Communication predicted higher scores on the task, while higher ratings of Leadership predicted lower scores. More clinically significant ratings of Conduct Problems predicted higher scores on Tower. Finally, Shift, Emotional Control, and Initiate predicted performance on Word Context with more clinically significant ratings of Shift predicting lower scores and more clinically significant ratings of Emotional Control and Initiate predicting higher scores.

Overall, the Adaptive and Internalizing factors appeared to best predict performance on D-KEFS variables. These results suggest that adaptive functioning may be a key area to assess when conducting an EF assessment, particularly in the area of set-shifting, given most of the D-KEFS variables included in the study are purported to measure this aspect of EF (Miller, 2013). Furthermore, results suggest that emotional state has the potential to impact performance on cognitive tasks and thus should be examined in neuropsychological assessments. It is likely that the Internalizing factor predicted performance on the D-KEFS because this factor included measures of set-shifting, which appears to play a role in internalizing disorders, such as anxiety (Hunter & Sparrow, 2012). Although these factors predicted performance on the D-KEFS, their influence is not clear given that higher ratings of Adaptive and Internalizing variables predicted both stronger and weaker performance.

Linear regression analyses with the subtests of the NEPSY-2 revealed that Somatization, Social Skills, Leadership, Plan/Organize, and Organization of Materials predicted performance on Animal Sorting. More clinically significant ratings of Somatization and Plan/Organize predicted higher scores, higher ratings of Social Skills predicted lower scores, higher ratings of Leadership predicted higher scores, and more clinically significant ratings of Organization of Materials and Monitor predicted lower scores. Monitor predicted performance on Auditory Attention with more clinically significant ratings predicting lower scores on this task. Anxiety and Somatization predicted performance on Response Set. While more clinically significant ratings of Anxiety predicted higher scores on this task, more clinically significant ratings of Somatization predicted lower scores on this task. Depression and Communication predicted performance on Word List Interference. More clinically significant ratings of Depression and higher ratings of Communication predicted higher scores on this task. No behavioral measures or factors were found to predict performance on Inhibition (Switching).

Overall, the Adaptive, Internalizing, and Metacognition factors best predicted performance on the NEPSY-2 variables. As mentioned previously, these results reinforce the notion that adaptive and emotional components should be assessed when addressing EF concerns through assessment as they appear to influence performance on EF-related tasks. Additionally, the Metacognition factor predicted performance on certain subtests of the NEPSY-2. While the D-KEFS measures included in this study primarily assess set-shifting, the NEPSY-2 variables assess other aspects of EF, including concept generation, sustained attention, and working memory, according to Miller (2013). These results suggest that the variables making up the Metacognition factor may be used to predict performance on various EF subcomponents with the exception of set-shifting, which is more accurately predicted by the variables making up the Shifting factor.

Lastly, behavioral variables were generally poor predictors of performance on the subtests of the WJ III COG NU. Communication predicted performance on Numbers Reversed with higher ratings of Communication predicting higher scores on Numbers Reversed. The overall Adaptive factor also predicted performance on this subtest. This result has multiple potential implications. First, it suggests that behavioral ratings of EF provide little insight into performance on measures of reasoning, including Concept Formation and Analysis/Synthesis. Second, communication skills likely predict performance on Numbers Reversed because this subtest requires mental rehearsal of verbal information as well as a verbal response. Lastly, this finding again reinforces the notion that adaptive functioning is a key area to assess in any evaluation because it has the potential to impact abilities in multiple areas of cognitive functioning.

Conclusions

The results of this study did not coincide with anticipated findings. First, it was anticipated that a two factor solution for behavioral measures of EF would be obtained, while a six factor solution for cognitive measures of EF would be obtained. In contrast, to this hypothesis, a five factor solution was revealed for behavioral measures of EF, while a two factor solution was revealed for cognitive measures of EF. It is likely that the behavioral measures of EF resulted in a larger number of factors than the cognitive measures of EF because certain BRIEF measures assess aspects of EF that are not paralleled in cognitive measures of EF (Toplak et al., 2013). For example, the BRIEF assesses one's ability to organize his or her work area, which is not assessed in any formal cognitive assessments. In this way, behavioral rating scales have the potential to provide information about one's functioning that is not available through formal testing.

In regard to the factor analysis of behavioral measures of EF, it is likely that more than two factors emerged because of the inclusion of the BASC-2 with the BRIEF. The BASC-2 variables loaded on four factors: Externalizing, Self-Regulation, Adaptive, and ADHD. Three of these four factors (i. e., Externalizing, Adaptive, and ADHD) do not necessarily measure specific aspects of EF. The remaining two factors (i.e., Self-Regulation and Metacognition) align with previous studies revealing a two-factor solution for behavioral ratings of EF (Gioia et al., 2000). A finding to note in the first factor analysis was the convergence of Internalizing variables of the BASC-2 with Self-Regulation variables of the BRIEF. The emergence of this factor supports the notion that internalizing disorders, such as depression and anxiety involve deficits in EF, particularly in the areas of self-regulation of emotion and set-shifting.

The factor analysis of cognitive measures of EF only produced two distinct factors. This is inconsistent with a recent study which demonstrated a six factor solution using nineteen performance-based measures purported to assess EF. Testa, Bennett, and Ponsford (2012) found the following factor solution for cognitive measures of EF in adults: Prospective Working Memory, Set-Shifting and Interference Management, Task Analysis, Response Inhibition, Strategy Generation and Regulation, and Self-Monitoring and Set-Maintenance. The two factors identified in the current study coincide with two of the factors attained in the aforementioned study (i.e., Set-Shifting and Interference Management and Strategy Generation and Regulation). It is likely that the other factors did not emerge during the current study because there were very few variables included that are believed to measure other key aspects of EF, including attention, working memory, and inhibition. This was particularly true after the TEACH and select NEPSY-2 subtests were removed from the analyses due to large amounts of missing data and low factor loadings, respectively.

Another important component of the second factor analysis is that the NEPSY-2 scores were eliminated completely due to low factor loadings. These low factor loadings may have resulted from the previously mentioned concern that there were too few variables included in the analysis measuring certain aspects of EF, such as working memory and inhibition. However, Response Set and Inhibition (Switching) were both removed from the analysis although they are proposed to measure set-shifting (Miller, 2013). By nature, EF tasks have a large degree of task impurity because EF is dependent upon more basic processes, such as sensorimotor skills (Miyake et al., 2000). Task impurity may explain the divergence of the NEPSY-2 subtests from the other measures of set shifting. Furthermore, the D-KEFS shifting tasks are structured in similar ways and often provide more instruction than the NEPSY-2 shifting tasks. These differences in test structure and task demands may further explain the elimination of the NEPSY-2 subtests from the factor solution.

The second and third hypotheses anticipated that some of the behavioral and cognitive factors obtained during the first phase of analysis would correlate with one another and that the variables comprising the behavioral factors would predict performance on the variables making up the related cognitive factors. Again, the results of this study did not align with hypotheses. Although research has varied on this subject, numerous studies have demonstrated poor associations between behavioral and cognitive measures of EF. For example, Bodnar, Prahme, Cutting, Denckla, and Mahone (2007) found no relationship between performance on the D-KEFS and behavioral ratings of EF. Additionally, Toplak et al. (2009) found that behavioral rating scales demonstrated diagnostic utility in predicting ADHD but showed very little overlap with performance-based measures of EF. These findings suggest that behavioral rating scales provide meaningful information that is different than the information provided by performance-based measures.

Furthermore, Toplak et al. (2013) analyzed twenty studies examining correlations between behavioral ratings of EF and performance-based measures of EF. Of the 286 correlations conducted within these studies, only 68 were significant. The authors argue that behavioral and cognitive measures of EF assess different aspects of EF by capturing information about separate levels of cognition. Specifically, the levels assessed by behavioral and cognitive measures of EF are termed reflective and algorithmic minds, respectively. According to Stanovich (2011), performance-based measures require the use of information processing mechanisms, such as perception and working memory, within the brain. On the other hand, behavioral rating scales take into account a reflective level of functioning, assessing the influence of one's goals, beliefs, and personal choice on everyday behaviors.

Barkley (1997) further supports the notion that behavioral rating scales and performance-based measures assess different aspects of functioning while recognizing the importance of assessing both. In general, cognitive measures provide information based on performance in a brief period of time, impairing the ability to draw concrete conclusions about one's functioning outside of the test environment. Gioia, Isquith, and Kenealy (2008) also recognize that aspects of the test environment, such as the structured nature of tasks and cuing by the examiner, may artificially enhance one's EF skills by reducing demands on inhibition, flexibility, planning, and goal-direction. In other words, artificial test environments may result in optimal performance, allowing deficits to be easily overlooked. On the other hand, behavioral rating scales provide information regarding every day, day-to-day functioning and are therefore more ecologically valid than cognitive measures of EF.

Although cognitive measures are considered less ecologically valid than behavioral measures of EF, they are still very useful in their ability to distinguish between specific strengths and weaknesses by comparing one's performance to a norm group. The results of this study support the notion outlined by Barkley (1997) that behavioral and cognitive measures of EF should not be considered equivalent or interchangeable. Rather, they measure different aspects of EF and each provide key information. When conducting an assessment of EF capacities, it is important to complete a thorough evaluation that considers individual performance as well as information from caregivers and other personnel close to the child. In other words, behavioral rating scales do not provide a shortcut for assessing EF; instead, best practice in school neuropsychology involves an in depth examination of neurocognitive processes using information from multiple sources as described by Miller (2013). Behavioral rating scales may be used as a screener for EF deficits but do not adequately assess all aspects of EF and therefore should not be used alone.

Limitations

The results of the current study are somewhat limited by a number of various factors, including use of archival data, characteristics of the sample, statistical techniques, and use of outdated test instruments. However, multiple safeguards and thoughtful decisions were made during the design of this study to ensure that the negative impact of these concerns was reduced.

Archival Data

There are several limitations associated with the use of archival data. Firstly, the initial collection of data was unable to be monitored. Therefore, it is not guaranteed that assessment instruments were administered according to standardization guidelines, possibly affecting the reliability and validity of scores. It is expected that some degree of consistency in administration and scoring occurred across practitioners because each received supervision from a practitioner belonging to the KIDS, Inc training program. Data entry errors may have occurred due to the large size of the dataset and variety of variables included within it. However, to best prevent entry errors, graduate students who entered data were trained in the entry and management of the dataset and worked in teams to enter data and check one another's work.

Additionally, because the data was collected based on individual referral questions, test batteries within the sample varied significantly, limiting the availability of data for certain assessments. As a result, 23.42% of the data in the current study was found to be missing completely at random. This amount of missing data reduced the power of the analyses and restricted the number of variables available to be included in the analysis. Ideally, behavioral measures of EF in addition to the BRIEF would have been included in the analysis due to concerns regarding the validity of the BRIEF. Furthermore, additional performance-based measures of EF, particularly in the areas of working memory and attention, may have increased the number of factors extracted during the second factor analysis.

Characteristics of the Sample

Certain characteristics of the sample further limit the interpretability of the current study. Although the use of a mixed-clinical sample is beneficial in investigating various diagnoses, the classifications of children included in the current study cannot be confirmed, as they were made by numerous practitioners. Additionally, the dataset included cases with primary, secondary, and tertiary diagnoses; therefore, many of the cases included were very complex. Each practitioner was trained and licensed to evaluate children but differences in practice may have resulted in variability and inaccuracy. Furthermore, the use of a mixed-clinical sample also reduces the generalizability of results (i.e., external validity) to the general population of children who do not belong to a clinical group (Gravetter & Forzano, 2012). Generalizability is further reduced because the majority of the sample was comprised of Caucasian males. Finally, the age range was reduced to ages 8 to 16 because of differences in age ranges of subtests selected for use in the current study. Therefore, results will not be generalizable to a younger age group when critical EF development is known to occur (Barkley, 2012b).

Despite the aforementioned concerns, the results of this study have the potential to inform the general understanding of the assessment of various EF components in children with clinical diagnoses. This study was not aimed at examining EF measures in children with specific types of diagnoses, but rather, in children with a variety of disorders that may involve EF deficits. Therefore, the results of this study should be interpreted carefully, recognizing that they apply only to children between the ages of eight and sixteen years who have been diagnosed with a clinical disorder, such as ADHD, autism spectrum disorder, a learning disability, traumatic brain injury, or others.

Statistical Techniques

The first limitation related to statistical techniques involves the use of full information maximum likelihood (FIML) to account for missing data during the factor analysis phase of the study. By nature, this technique is flawed because it uses statistical procedures to attempt to predict what the most likely data points would be to replace missing data. Therefore, this study is limited because it uses these data points for 23.42% of the data, rather than true data points. It was determined that this procedure was adequate for the current study given findings that FIML estimates are more efficient and less biased than other methods (i.e., listwise deletion, pairwise deletion, and response pattern imputation) for accounting for data missing completely at random (Enders & Bandalos, 2001). Furthermore, FIML is best utilized when sample sizes are larger than 100 like the sample used as part of the current study (Jain & Wang, 2008).

The use of exploratory factor analysis further limits the current study. This method is considered exploratory in nature, meaning it is often the first step taken to examine factor structures, followed by more complex procedures that test the fit of models. Exploratory factor analysis is also strongly influenced by the decisions of the researcher and therefore can produce different results based on decisions made during each factor extraction phase. To address this issue, specific guidelines were followed to determine which variables to remove from the analysis. Specifically, variables were removed from each analysis one at a time. Factors with the lowest factor ratings that fell below the cutoff of .5 or those loading on multiple factors were removed from the analysis individually. Finally, factors obtained during factor analysis require researchers to interpret and name each factor. Decisions for naming each of the factors were made based upon well-known models of EF (Miller, 2013; Barkley, 2012b).

Test Instruments

A final concern related to the current study results from the use of somewhat outdated versions of the BASC-2, BRIEF, and WJ III COG NU. New versions of each of these assessment tools are now available. However, these assessment tools were considered state of the art tests at the time they were administered; additionally, they did not undergo significant changes between the most recent versions. They include many of the same questions and subtests with very slight or no variations with the exception of updated normative data. Because the structure of these tests has not changed, the new versions of the BASC and BRIEF continue to provide scores for each of the variables included in the current study and are purported to measure the same behaviors as the previous versions. Furthermore, the WJ III COG NU subtests included in the current study are almost identical to the corresponding subtests in the new version of the test and measure the same aspects of cognitive functioning according to CHC theory (Flanagan & Harrison, 2012). Finally, it is important to note the use of older versions of the tests allows for increased statistical power because more data is available for these versions of the tests. Therefore, it was determined that the results of the current study would provide valuable information despite use of outdated instruments.

Future Research

The natural follow-up to this study would be to investigate the obtained factor structures for the behavioral and cognitive measures of EF through confirmatory factor analysis. However, because the obtained models did not explain an adequate amount of variance, it would not be logical to test the fit of these models. Instead, it would be useful to complete a similar study analyzing behavioral and cognitive variables together in one large exploratory factor analysis. In order to obtain an adequate amount of power for so many variables, the sample size would have to be at least twice as large as the sample size of the current study. Additionally, other measures of EF should be included in the analysis, particularly those measuring working memory and aspects of attention, in order to increase the likelihood of subtests loading on factors and of obtaining a larger number of EF factors as found by Testa et al. (2012). This study could be completed with both parent and teacher ratings as it would be useful to compare results given that parents and teachers observe students in very different environments.

Given the complex developmental trajectory of EF, another area of future research would be to examine the factor structures of measures of EF in different age groups. Specifically, less complex EF skills, such as attention, often develop at younger ages, while more complex EF skills, such as goal-directed thought and reasoning develop during adolescence (Zelazo et al., 2003). Based on this developmental trend, it would be anticipated that more EF factors would be extracted when examining measures of EF in older children than in younger children. The current study included participants between the ages of eight and sixteen, which may explain why factors involving complex components of EF were extracted but measures involving less complex components were not included in the final factor solutions. In other words, because complex EF skills rely heavily on more basic processes, such as attention, it makes sense that factors representing attention and other basic skills did not emerge, particularly if all other variables required the use of attention.

Finally, the finding that internalizing symptoms as measured by the BASC-2 loaded on the same factor as self-regulation measures of the BRIEF, suggests that there is some overlap between anxiety, depression, somatization, set-shifting, and emotional regulation. Fossati, Ergis, and Allilaire (2002) provide an overview of the EF impairments often observed in patients with depressive disorders. Individuals with depression often exhibit impairments in cognitive inhibition, leading to difficulty filtering out information and controlling mood changes. Additionally, poor cognitive flexibility associated with depression has been linked to perpetuation of depressive states while deficits in planning have been linked to low motivation. Anxiety, attentional control, including inhibition and shifting, have been associated with difficulty shifting attention away from threatening stimuli and with obsessive-compulsive symptoms, such as intrusive thoughts (Eysenck, Derakshan, Santos, & Calvo, 2007). It is clear that there is some degree of relationship between EF and internalizing disorders; however, more research is needed to clarify the role of executive function in internalizing disorders. Specifically, it is unclear whether deficits in cognitive inhibition, problem-solving, planning, and shifting result in internalizing symptoms, such as rumination or intrusive thoughts, or whether anxious or depressive states result in associated EF impairments. Additionally, a study investigating the factor structure of EF in samples of individuals with anxiety and depression may provide more information regarding the nature of EF in these clinical populations.

Final Thoughts

Although the construct of EF has been researched extensively in recent years, there remains substantial disagreement about the definition of EF and about the components that make up the construct (Eslinger, 1996; Packwood et al., 2011). Because of this, a clear, concise, and testable theory of EF does not exist. Instead, multiple theories outline various proposed structures of EF, which typically include, attentional control, planning, organization, problem-solving, and behavioral regulation (Hunter & Sparrow, 2012; McCloskey et al., 2009). The current study supports the notion that cognitive flexibility or set shifting, emotional regulation, reasoning, and metacognitive skills are key features of EF measured by behavioral rating scales and performance-based measures of EF. However, more importantly, the results of this study provide additional evidence that the construct of EF is highly complex, making it difficult to define and effectively measure. A key reason for this complexity is the nature of frontal lobe functions associated with EF, which recruit more basic processes within the brain in order to engage in higher-order thought (Hunter & Sparrow, 2012). In this way, EF involves a complex system of neural networks and structures within the brain. Just as it does not rely on functioning in one specific area of the brain, it does not involve a clear pattern of behavioral or cognitive deficits. Everyday behaviors requiring the regulation of EF and performance on cognitive based measures of EF are influenced by numerous individual factors, which interact in unique ways. When attempting to measure EF, these additional factors affect findings, resulting in task impurity of EF measures.

Previous factor analytic studies resulted in similar conclusions as those discussed in the context of the current study. Testa et al. (2012) analyzed numerous studies exploring the factor structure of EF and concluded that EF is an extremely complex construct that is difficult to define and measure because of significant variability within individual EF systems. In a precursor study, Avirett (2011) conducted a confirmatory factor analysis utilizing selected NEPSY-2, D-KEFS, and WJ III COG subtests to examine the factor structure of EF. Results indicated that the factor structure of EF is both difficult to discern and extremely variable. Difficulty identifying a clear factor structure most likely relates to the complex nature of EF itself as well as the breadth of tasks purported to measure various aspects of EF (Packwood, 2001). Because theorists and researchers disagree on the best definitions of and methods of measuring EF, there is significant variability between measurement tools, making it difficult for them to converge in a specific manner.

Regarding the relationship between behavioral ratings of EF and performance-based measures of EF, the results of this study support Barkley's statements made over two decades ago (1997). That is, behavioral ratings and scores on cognitive measures of EF should not be interpreted as equivalent or interchangeable. These measures should both be considered key components of neuropsychological assessments, including those investigating EF, because they measure different aspects of the construct. Whereas performance-based measures are time-limited and constrained by specific task demands, behavioral ratings provide ecologically valid information about a child's day-to-day functioning. It is important to consider both a child's strengths and weaknesses in a highly structured environment as well as his or her behavior as it is affected by outside distractions and various environments. In other words, the assessment of EF should be considered equally complex as the construct itself.

REFERENCES

Abdi, H. (2007). RV coefficient and congruence coefficient, In N. J. Salkind (Ed.), *Encyclopedia of measurement and statistics.* Thousand Oaks: Sage.

Abelson, R. P. (1963). Computer simulation of "hot cognitions." In S. Tomkins & S. Messick (Eds.), *Computer simulation of personality* (pp. 277-298). New York, NY: Wiley.

Achenbach, T. (1991a). *Manual for the Child Behavior Checklist and 1991 profile.* Burlington: University of Vermont, Department of Psychiatry.

Achenbach, T. (1991b). *Manual for the Teacher's Report Form and 1991 profile.* Burlington: University of Vermont, Department of Psychiatry.

Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA School-Age Forms & Profiles.* Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.

Adler, N. E., & Rehkopf, D. H. (2008). U. S. disparities in health: Descriptions, causes, and mechanisms. *Annual Review of Public Health, 29*, 235-252.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5^{th} ed.). Arlington, VA: American Psychiatric Publishing.

Anderson, P. J. (2002). Assessment and development of executive function (EF) during childhood. *Child Neuropsychology, 8*, 71-82.

Anderson, P. J. (2008). Towards a developmental model of executive function. In V. Anderson, R. Jacobs, & P. J. Anderson (Eds.), *Executive functions and the frontal lobes: A lifespan perspective* (pp. 3-21). New York, NY: Psychology Press.

Anderson, V. A., Anderson, P., Northam, E., Jacobs, R., & Catroppa, C. (2001). Development of executive functions through late childhood and adolescence in an Australian sample. *Developmental Neuropsychology, 20*, 385-406.

Anderson, V., Catroppa, C., Morse, S., Haritou, F., & Rosenfeld, J. (2000). Recovery of intellectual ability following traumatic brain injury in childhood: Impact of injury severity and age at injury. *Pediatric Neurosurgery, 32*, 282-290.

Ardila, A. (2008). On the evolutionary origins of executive functions. *Brain and Cognition, 68*, 92-99.

Arizmendi, T., Paulson, K., & Domino, G. (1981). The Matching Familiar Figures Test: A primary, secondary, and tertiary evaluation. *Journal of Clinical Psychology, 37*, 812-818.

Aron, A. R. (2008). Progress in executive-function research: From tasks to functions to regions to networks. *Current Directions in Psychological Science, 17*, 124-129.

Aron, A. R., Behrens, T. E., Smith, S., Frank, M. J., & Poldrack, R. A. (2007). Triangulating a cognitive control network using diffusion-weighted magnetic resonance imaging (MRI) and functional MRI. *Journal of Neuroscience, 27*, 3743-3652.

Avirett, E. (2011). *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.* (Doctoral Dissertation) Texas Woman's University, Denton, TX. Retrieved from Proquest Dissertations and Theses Database. (UMI No. 3464570).

Baddeley, A. D. (1986). Modularity, mass-action and memory. *The Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 38*, 527-533.

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*, 128-137.

Baird, A. A., Kagan, J., Gaudette, T., Walz, K. A., Hershlag, N., & Boas, D. A. (2002). Frontal lobe activation during object permanence: Data from near-infrared spectroscopy. *NeuroImage, 16*, 1120-1126.

Banich, M. T. (2009). Executive function: The search for an integrated account. *Current Directions in Psychological Science, 18*, 89-94.

Baldo, J. V., Shimarmura, A. P., Delis, D. C., Kramer, J., & Kaplan, E. (2001). Verbal and design fluency in patients with frontal lobe lesions. *Journal of International Neuropsychochological Society, 7*, 586-596.

Barkley, R. A. (1997). Behavioral inhibition, sustained attention, and executive functions: Constructing a unifying theory of ADHD. *Psychological Bulletin, 121*, 65-94.

Barkley, R. A. (2012a). *Barkley Deficits in Executive Functioning Scale: Children and Adolescents (BDEFS-CA).* New York, NY: The Guildford Press.

Barkley, R. A. (2012b). Executive functions: What they are, how they work, and why they evolved. New York, NY: The Guilford Press.

Baune, B. T., Konrad, C., Suslow, T. Domschke, K., Birosova, E., Sehlmeyer, C., & Beste, C. (2010). The reelin (RELN) gene is associated with executive function in healthy individuals. *Neurobiology of Learning and Memory, 94*, 446-451.

Bayliss, D. M., & Roodenrys, S. (2000). Executive processing and attention deficit hyperactivity disorder: An application of the supervisory attentional system. *Developmental Neuropsychology, 17*, 161-180.

Belloni, K. C. (2012). A confirmatory factor analytic comparison of the Test of Everyday Attention for Children. *Dissertation Abstracts International: Section A. Humanities and Social Sciences, 72*(10), 3610.

Berg, E. A. (1948). A simple objective test for measuring flexibility in thinking. *Journal of General Psychology, 39*, 15-22.

Best, J. R., Miller, P. H., & Jones, L. L. (2009). Executive functions after age 5: Changes and correlates. *Developmental Review, 29*, 180-200.

Bierman, K. L., Nix, R. L., Greenberg, M. T., Blair, C., & Domitrovich, C. E. (2008). Executive functions and school readiness intervention: Impact, moderation, and mediation in the Head Start REDI program. *Development and Psychopathology, 20*, 821-843.

Binet, A. A., Simon, T. T., & Town, C. H. (1912). *A method of measuring the development of the intelligence of young children.* Lincoln, IL: Courier Company.

Blakemore, S., & Choudhury, S. (2006). Development of the adolescent brain: Implications for executive function and social cognition. *Journal of Child Psychology and Psychiatry, 47*, 296-312.

Blair, C., & Diamond, A. (2008). Biological processes in prevention and intervention: The promotion of self-regulation as a means of preventing school failure. *Development and Psychopathology, 20*, 899-911.

Blair, C., & Ursache, A. (2011). A bidirectional model of executive functions and self-regulation. In K. D. Vohs & R. F. Baumeister (Eds.), *Handbook of self-regulation: Research, theory, and applications* (2nd ed., pp. 300-320). New York, NY: Guilford Press.

Blumenfeld, H. (2010). *Neuroanatomy through clinical cases* (2nd ed.). Sunderland, MA: Sinauer Associates.

Bodnar, L. E., Prahme, M. C., Cutting, L. E., Denckla, M. B., Mahone, E. M. (2007). Construct validity of parent ratings of inhibitory control. *Child Neuropsychology, 13*, 345-362.

Bradley, R. H., & Corwyn, R. F. (2002). Socioeconomic status and child development. *Annual Review of Psychology, 53*, 371-399.

Brocki, K., Fan. J., & Fossella, J. (2008). Placing neuroanatomical models of executive function in a developmental context: Imaging and imaging-genetic strategies. *New York Academy of Sciences, 1129*, 246-255.

Brown, T. E. (2001). *Brown attention-deficit disorder scales for children and adolescents.* San Antonio: The Psychological Corporation.

Burgess, P. W., Alderman, N., Evans, J., Emslie, H., & Wilson, B. (1998). The ecological validity of tests of executive function. *Journal of the International Neuropsychological Society, 4*, 547-558.

Burt, C. (1948). Factor analysis and canonical correlations. *British Journal of Psychology, Statistical Section, 1*, 95-106.

Carlson, S. M. (2003). Executive function in context: Development, measurement, theory, and experience. *Monographs of the Society for Research in Child Development, 68*, 138-151.

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

Casey, B. J., Trainor, R. J., Orendi, J. L., Schubert, A. B., Nystrom, L. N., Giedd, J. N., . . . Rapoport, J. L. (1997). A developmental functional MRI study of prefrontal activation during a go-no go task. *Journal of Cognitive Neuroscience, 9*, 835-847.

Champagne, F. A. (2010). Early adversity and developmental outcomes: Interaction between genetics, epigenetics, and social experiences across the lifespan. *Perspectives on Psychological Science, 5,* 564-574.

Chan, R. C. K., Wang, L., Ye, J., Leung, W. W. Y., & Mok, M. Y. K. (2008). A psychometric study of the Test of Everyday Attention for Children in the Chinese setting. *Archives of Clinical Neuropsycholgy, 23*, 455-466.

Chudasama, Y. (2011). Animal models of prefrontal-executive function. *Behavioral Neuroscience, 125*, 327-343.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrene Erlbaum Associates.

Cohen, M. J. (1997). *Children's Memory Scale: Examiner's manual.* San Antonio:

Harcourt Brace & Company.

Collins, L. M., Schafer, J. L., & Kame, C. M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods, 6,* 330-351.

Comrey A. L., & Lee, H. B. (1992). *A first course in factor analysis.* Hillsdale, NJ: Erlbaum.

Conners, C. K. (1989). *Manual for Conners' Rating Scales.* North Towanda, NY: Multi-Health Systems.

Conners, C. K. (1997). *The Conners Rating Scales – Revised manual.* North Towanda, NY: Multi-health Systems.

Conners, C. K. (2013). *Conners' CPT 3: Conners' Continuous Performance Test Third Edition* [computer software]. Toronto: Multi Health Systems.

Cripe, L. I. (1996). The ecological validity of executive function testing. In R. J. Sbordone, & C. J. Long (Eds.), *Ecological validity of neuropsychological testing* (pp. 171-202). Delray Beach, FL: GR Press/St. Lucie Press.

Crone, E. A. (2009). Executive functions in adolescence: Inferences from brain and behavior. *Developmental Science Review, 12*, 825-830.

Davis, A. S. (2011). *Handbook of pediatric neuropsychology.* New York, NY: Springer Publishing.

Dawson, P., & Guare, R. (2010). *Executive skills in children and adolescents: A practical guide to assessment and intervention* (2nd ed.). New York, NY: The Guilford Press.

Dehn, M. J. (2008). *Working memory and academic learning: Assessment and intervention.* Hoboken, NJ: John Wiley & Sons, Inc.

De Jong, P. F. (1998). Working memory deficits of reading disabled children. *Journal of Experimental Child Psychology, 70*, 75-96.

Delis, D. C. (2012). *Delis Rating of Executive Functions (D-REF).* Bloomington, MN: Pearson, Inc.

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

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

Denckla, M. B. (1996). A theory and model of executive function: A neuropsychological perspective. In G. R. Lyon & N. A. Krasnegor (Eds.), *Attention, memory, and executive function* (pp. 263-278). Baltimore, MD: Paul H. Brookes.

Dennis, M., Guger, S., Roncadin, C., Barnes, M., & Schachar, R. (2001). Attentional-inhibitory control and social-behavioral regulation after childhood closed head injury: Do biological, developmental, and recovery variables predict outcome? *Journal of the International Neuropsychological Society, 7*, 683-692.

Diaz, R. M., & Berk, L. E. (1992). *Private speech: From social interaction to self-regulation.* Hillsdale, NJ: Lawrence Erlbaum.

Donders, J. (2002). The behavior rating inventory of executive function: Introduction. *Child Neuropsychology, 8,* 229-230.

DuPaul, G. J., Power, T. J., Anastopoulos, A. D., & Reid, R. (1998). *ADHD Rating*

Scale-IV: Checklist, norms, and clinical interpretation. New York: Guilford Press.

Elliott, C. D. (1990). *Differential Ability Scales.* San Antonio: Psychological Corporation.

Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal, 8*, 430-457.

Eslinger, P. J. (1996). Conceptualizing, describing, and measuring components of executive function: A summary. In G. R. Lyon & N. A. Krasnegor (Eds.), *Attention, memory, and executive function* (pp. 367-395). Baltimore, MD: Paul H. Brookes.

Eslinger, P. J., & Damasio, A. R. (1985). Severe disturbance of higher cognition after bilateral frontal lobe ablation: Patient EVR. *Neurology, 35*, 1731-1741.

Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: Attentional control theory. *Emotion, 7*, 336-353.

Ferrando, P. J., Lorenzo, U. (2000). Unrestricted versus restricted factor analysis of multidimensional test items: some aspects of the problem and some suggestions. *Psychologica, 21*, 301-323.

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). Thousand Oaks, CA: SAGE.

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

Fossati, P., Ergis, A. M., & Allilaire, J. F. (2002). Executive functioning in unipolar depression: A review. *Encephale, 28*, 97-107.

Franzen, M. D., & Wilhelm, K. L. (1996). Conceptual foundation of ecological validity in neuropsychological assessment. In Sbordone, R. J. & Long, C. J (Eds.), *Ecological Validity of Neuropsychological Testing*. Delray Beach, FL: St. Lucie Press.

Friedman, N. P., Miyake, A., Young, S. E., DeFries, J. C., Corley, R. P., & Hewitt, J. K. (2008). Individual differences in executive function are almost entirely genetic in origin. *Journal of Experimental Psychology: General, 137*, 201-225.

Frith, U., & Happe, F. (1994). Autism: Beyond "theory of mind." *Cognition, 50*, 115-132.

Fuster, J. M. (1997). *The prefrontal cortex.* New York: Raven.

Ganesalingam, K., Yeates, K. O., Taylor, H. G., Walz, N. C., Stancin, T., & Wade, S. (2011). Executive functions and social competence in young children 6 months following traumatic brain injury. *Neuropsychology, 25*, 466-476.

Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Bulletin, 134*, 31-60.

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

Gilotty, L., Kenworthy, L., Sirian, L., Black, D. O., & Wagner, A. E. (2002). Adaptive skills and executive function in autism spectrum disorders. *Child Neuropsychology, 8*, 241-248.

Gioia, G. A., & Isquith, P. K. (2004). Ecological assessment of executive function in traumatic brain injury. *Developmental Neuropsychology, 25,* 135-158.

Gioia, G. A., Isquith, P. K., Guy, S. C., & Kenworthy, L. (2000). *Behavior Rating Inventory of Executive Function*. Lutz, FL: Psychological Assessment Resources.

Gioia, G. A., Isquith, P. K., Guy, S. C., & Kenworthy, L. (2015). *Behavior Rating Inventory of Executive Function, Second Edition (BRIEF-2).* Lutz, FL: Psychological Assessment Resources.

Gioia, G. A., Isquith, P. K., Kenworthy, L., & Barton, R. M. (2002). Profiles of everyday executive function in acquired and developmental disorders. *Child Neuropsychology, 8,* 121-137.

Goldberg., E. (2001). *The executive brain: Frontal lobes and the civilized mind.* New York, NY: Oxford University Press.

Golden, C. J., & Freshwater, S. M. (2001). Luria-Nebraska Neuropsychological Battery. In W. I. Dorfman & M. Hersen (Eds.), *Understanding psychological assessment* (pp. 59-75). Dordrecht, Netherlands: Kluwer Academic Publishers.

Golden, C. J., & Freshwater, S. M. (2002). *Stroop Color and Word Test: A manual for clinical and experimental uses.* Chicago: Stoelting.

Gorsuch, R. L. (1997). Exploratory factor analysis: its role in item analysis. *Journal of Personality Assessment, 68*, 532-560.

Gorton, D. (1987). Gestalt therapy: The historical influences on Frederick S. Perls. *Gestalt Theory, 9*, 28-39.

Grant, D. A., & Berg, E. (1948). A behavioral analysis of degree of reinforcement and ease of shifting to new responses in Weigl-type card-sorting problem. *Journal of*

Experimental Psychology, 38, 404-411.

Gravetter, F. J., & Forzano, L. B. (2012). *Research methods for the behavioral sciences,* (4th ed.). Independence, KY: Cengage.

Guthrie, R. V. (2004). *Even the rat was white: A historical view of psychology* (2nd ed.). Boston, MA: Pearson.

Hackman, D. A., Farah, M. J., & Meaney, M. J. (2010). Socioeconomic status and the brain: Mechanistic insights from human and animal research. *Nature Reviews Neuroscience, 11*, 651-659.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate Data Anlaysis* (7th ed.). New York, NY: Pearson.

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

Halperin, J. M., & Healey, D. M. (2011). The influences of environmental enrichment, cognitive enhancement, and physical exercise on brain development: Can we alter the developmental trajectory of ADHD? *Neuroscience and Biobehavioral Reviews, 35*, 621-634.

Harlow, J. M. (1848). Recovery from the passage of an iron bar through the head. *Boston Medical and Surgical Journal, 39*, 389-393.

Heaton, R. K., Chelune, G. J., Talley, J. L., Kay, G. G., & Curiss, G. (1993). *Wisconsin Card Sorting Test manual: Revised and expanded.* Odessa, FL: Psychological Assessment Resources, Inc.

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized intelligence. *Journal of Educational Psychology, 57*, 253-270.

Hosenbocus, S., & Chalal, R. (2012). A review of executive function deficits and pharmacological management in children and adolescents. *Journal of the Canadian Academy of Child & Adolescent Psychiatry, 21*, 223-229.

Hughes, C. H. (2011). Changes and challenges in 20 years of research into the development of executive functions. *Infant and Child Development, 20*, 251-271.

Hughes, C. H., & Ensor, R. A. (2009). How do families help or hinder the emergence of early executive function? *New Directions in Child and Adolescent Development, 123*, 35-50.

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

Hughes, C., Ensor, R., Wilson, A., & Graham, A. (2010). Tracking executive function across the transition to school: A latent variable approach. *Developmental Neuropsychology, 35*, 20-36.

Hughes, C., Roman, G., Hart, M. J., & Ensor, R. (2012). Does maternal depression predict young children's executive function? A 4-year longitudinal study. *Journal of Child Psychology and Psychiatry, 54*, 169-177.

Hunter, S. J., & Sparrow, E. P. (2012). *Executive function and dysfunction: Identification, assessment, and treatment.* New York, NY: Cambridge University Press.

Huttenlocher, P. R. (1979). Synaptic density in human frontal cortex: Developmental changes and effects of aging. *Brain Research, 163*, 195-205.

Jain, R. B., & Wang, R. Y. (2008). Limitations of maximum likelihood estimation procedures when a majority of the observations are below the limit of detection. *Analytic Chemistry, 80*, 4767-4772.

Jensen, A. R. (1998). *The g factor: The science of mental ability.* Westport, CT: Praeger.

Jurado, M. B., & Rosselli, M. (2007). The elusive nature of executive functions: A review of our current understanding. *Neuropsychological Review, 17*, 213-233.

Justak, J. F., & Wilkinson, G. S. (1984). *Wide Range Achievement Test-Revised.* Washington: Justak Assessment Systems.

Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika, 23*, 187-200.

Kaitaro, T. (2001). Biological and epistemological models of localization in the nineteenth century from Gall to Charcot. *Journal of the History of the Neurosciences, 10*, 262-276.

Kaplan, E. (1988). The process approach to neuropsychological assessment. *Aphasiology, 2*, 309-311.

Kaufman, A. S., & Kaufman, N. L. (1993). Kaufman Adolescent and Adult Intelligence Test (KAIT). Circle Pines, MN: American Guidance Service.

Keilp, J. G., Wyatt, G., Gorlyn, M., Oquendo, M. A., Burke, A. K., & Mann, J. J. (2014). Intact alternation performance in high lethality suicide attempters. *Psychiatry Research, 219*, 129-136.

Kishiyama, M. M., Boyce, W. T., Jimenez, A. M., Perry, L. M., & Knight, R. T. (2009). Socioeconomic disparities affect prefrontal function in children. *Journal of*

Cognitive Neuroscience, 21, 1106-1115.

Korkman, M., Kirk, U., & Kemp, S. (2007). *NEPSY-II: Clinical and interpretive manual.* San Antonio: The Psychological Corporation.

Lambek, R., Tannock, R., Dalsgaard, S., Trillingsgaard, A., Damm, D., & Thomsen, P. H. (2010). Validating neuropsychological subtypes of ADHD: How do children with and without an executive function deficit differ? *Journal of Child Psychology and Psychiatry, 51*, 895-904.

Leve, L. D., DeGarmo, D. S., Bridgett, D. J., Neiderhiser, J. M., Shaw, D. S., Harold, G. T., . . . Reiss, D. (2013). Using an adoption design to separate genetic, prenatal, and temperament influences on toddler executive function. *Developmental Psychology, 49,* 1045-1057.

Levin, B. E. (1990). Organizational deficits in dyslexia: Possible frontal lobe dysfunction. *Developmental Neuropsychology, 6*, 95-110.

Levin, H. S., Culhane, K. A., Mendelsohn, E., Lilly, M. A., Bruce, D., Fletcher, J., . . . Eisenberg, H. M. (1993). Cognition in relation to magnetic resonance imaging in head-injured children and adolescents. *Archives of Neurology, 50*, 897-905.

Lezak, M. D., (1982). The problem of assessing executive functions. *International Journal of Psychology, 17,* 281-297.

Lezak, M. D. (1995). *Neuropsychological Assessment.* New York: Oxford University Press.

Lezak, M. D. (2004). *Neuropsychological assessment* (4th ed.). New York: Oxford University Press.

Lipina, S. J., Martelli, M. I., Vuelta, B., & Colombo, J. A. (2005). Performance on the A-not-B task of Argentinean infants from unsatisfied and satisfied basic needs homes. *International Journal of Psychology, 39*, 49-60.

Liss, M., Fein, D., Allen, D., Dunn, M., Feinstein, C., Morris, R., . . . Rapin, I. (2001). Executive functioning in high-functioning children with autism. *Journal of Child Psychology, 42*, 261-270.

Little, R. J., & Rubin, D. B. (2002). Statistical analysis with missing data. Hoboken: John Wiley & Sons, Inc.

Logue, S. F., & Gould, T. J. (2014). The neural and genetic basis of executive function: Attention, cognitive flexibility, and response inhibition. *Pharmacology, Biochemistry and Behavior, 123,* 45-54.

Lorenzo-Seva, U. & ten Berge, J. M. F. (2006). Tucker's congruence coefficient as a meaningful index of factor similarity. *Methodology, 2*(2), 557-64.

Luria A. R. (1966). *Higher cortical functions in man.* New York: Basic Books, Inc.

Luria, A. R. (1973). The frontal lobes and the regulation of behavior. In K. H. Pribam & A. R. Luria (Eds.), *Psychophysiology of the frontal lobes* (pp. 3-26). New York: Academic Press.

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

Manchester, D., Priestley, N., & Jackson, H. (2004). The assessment of executive functions: Coming out of the office. *Brain Injury, 18*, 1067-1081.

Manly, T., Anderson, V., Crawford, J., George, M., Underbjerg, M., & Robertson, I. H.

(2016). *The Test of Everyday Attention for Children, Second Edition (TEA-Ch2).* Oxford: Pearson.

Manly, T., Robertson, I. H., Anderson, V., & Nimmo-Smith, I. (1999). *The Test of Everyday Attention for Children (TEA-Ch)*. London: Harcourt Assessment.

Mashour, G. A., Walker, E. E., & Martuza, R. L. (2005). Psychosurgery: Past, present, and future. *Brain Research Review, 48*, 409-419.

Maricle, D. E., Johnson, W., & Averitt, E. (2010). Assessing and intervening in children with executive function disorders. In D. C. Miller (Ed.), *Best practices in school neuropsychology* (pp. 599-640). Hoboken, NJ: Wiley.

Mattson, S. N., Goodman, A. M., Caine, C., Delis, D. C., & Riley, E. P. (1999). Executive functioning in children with heavy prenatal alcohol exposure. *Alcoholism: Clinical and Experimental Research, 23*, 1808-1815.

McCloskey, G., Perkins, L. A., & Van Divner, B. (2009). *Assessment and intervention for executive function difficulties.* New York, NY: Routledge Taylor & Francis Group.

McCloskey, G., & Perkins, L. A. (2012). *Essentials of executive function assessment*. A.S. Kaufman & N. L. Kaufman (Eds). Hoboken: Wiley.

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

McIntosh, P. (2003). Unpacking the invisible knapsack. In S. Plous (Ed.), *Understanding prejudice and discrimination* (pp. 191-196). New York, NY: McGraw-Hill.

Meaney, M. J., Szyf, M., & Seckl, J. R. (2007). Epigenetic mechanisms of perinatal programming of hypothalamic-pituitary-adrenal function and health. *Trends in Molecular Medicine, 13*, 269-277.

Menon, V., & Uddin, L. Q. (2010). Saliency, switching, attention and control: A network model of insula function. *Brain Structure and Function, 214*, 655-667.

Meyers, L. S., Gamst, G., & Guarino, G. (2006). *Applied multivariate research: Design and interpretation.* Thousand Oaks, CA: Sage Publications.

Miller, M., Nevado-Montenegro, A. J., & Hinshaw, S. P. (2012). Childhood executive function continues to predict outcomes in young adult females with and without childhood-diagnosed ADHD. *Journal of Abnormal Child Psychology, 40*, 657-668.

Miller, D. C. (2013). *Essentials of school neuropsychological assessment* (2nd ed.). Hoboken: John Wiley & Sons, Inc.

Minshew, N. J., Goldstein, G., & Siegel, D. J. (1997). Neuropsychologic functioning in autism: Profile of a complex information processing disorder. *Journal of the International Neuropsychological Society, 3*, 303-316.

Mischel, W., Shoda, Y., & Rodriguez, M. L. (1989). Delay of gratification in children. *Science, 244*, 933-938.

Molfese, V. J., Molfese, P. J., Molfese, D. L., Rudasill, K. M., Armstrong, N., & Starkey, G. (2010). Executive function skills of 6 to 8 year olds: Brain and behavioral evidence and implications for school achievement. *Contemporary Educational Psychology, 35*, 116-125.

Moller, P. (1996). Materialistic monism and negative feedback control. *PsycCritiques, 41*, 675-676.

Neath, I., & Suprenant, A. M. (2003). *Human memory* (2nd ed.). Belmont, CA: Wadsworth.

Newman, M. G., Castonguay, L. G., Borkovec, T. D., Fisher, A. J., & Nordberg, S. S. (2008). An open trial of integrative therapy for generalized anxiety disorder. *Psychotherapy: Theory, Research, Practice, Training, 45*, 135-147.

Nigg, J. T., & Casey, B. J. (2005). An integrative theory of attention-deficit/hyperactivity disorder based on the cognitive and affective neurosciences. *Development and Psychopathology, 17*, 765-806.

Noble, K. G., McCandliss, B. D., & Farah, M. J. (2007). Socioeconomic gradients predict individual differences in neurocognitive abilities. *Developmental Science, 10*, 464-480.

Norman, D., & Shallice, T. (1986). Attention to action: Willed and automatic control of behavior. In R. Davidson, R. Schwartz, & D. Shapiro (Eds.), *Consciousness and Self-Regulation: Advanced in Research and Theory IV*. New York: Plenum Press.

Nunnally, J. C. (1978). *Psychometric Theory* (2nd ed.). New York: McGraw-Hill.

Ozonoff, S., & Cathcart, K. (1998). Effectiveness of a home program intervention for young children with autism. *Journal of Autism and Developmental Disorders, 28*, 25-32.

Packwood, S., Hodgetts, H. M., & Tremblay, S. (2011). A multiperspective approach to the conceptualization of executive functions. *Journal of Clinical and*

Experimental Neuropsychology, 33, 456-470.

Partington, J. E., & Leiter, R. G. (1949). Partington's pathway test. *Psychological Service Center Bulletin, 1*, 9-20.

Pennington, B. F., & Ozonoff, S. (1996). Executive functions and developmental psychopathology. *Journal of Child Psychology and Psychiatry, 37*, 51-87.

Poletti, M. (2009). Adolescent brain development and executive functions: A prefrontal framework for developmental psychopathologies. *Clinical Neuropsychiatry, 6*, 155-163.

Posner, M. I. (1994). Attention: The mechanism of consciousness. *Proceedings of the National Academy of Sciences, 91*, 7398-7402.

Pribham, K. H. (1973). The primate frontal cortex: Executive of the brain. In K. H. Pribham and A. R. Luria (Eds.), *Psychophysiology of the frontal lobes* (pp. 293-314). New York: Academic Press.

Qui, M., Ye, Z., Li, Q, Liu, G., Xie, B., & Wang, J. (2011). Changes of brain structure and function in ADHD children. *Brain Topography, 24*, 243-252.

Rakic, P. (2009). Evolution of the neocortex: A perspective from developmental biology. *Nature Reviews Neuroscience, 10*, 724-735.

Rappoport, J. L., Castellanos, F., Gogate, N., Janson, K., Kohler, S., & Nelson, P. (2001). Imaging normal and abnormal brain development: New perspectives for child psychiatry. *Australian and New Zealand Journal of Psychiatry, 35*, 272-281.

Reise, S. P., Waller, N. G., & Comrey, A. L. (2000). Factor analysis and scale revision. *Psychological Assessment, 12*, 287-297.

Reitan, R. M., & Wolfson, D. (1993) The Halstead-Reitan Neuropsychological Test Battery: Theory and clinical interpretation (2nd ed.). Tucson, AZ: Neuropsychology Press.

Reynolds, A. J. (1994). Effects of preschool plus follow-on intervention for children at risk. *Developmental Psychology, 30*, 787-804.

Reynolds, C. R., & Kamphaus, R. W. (1992). *Behavior assessment system for children.* Circle Pines, MN: American Guidance Service.

Reynolds, C. R., & Kamphaus, R. W. (2004). *Behavior assessment system for children* (2nd ed.). Circle Pines, MN: American Guidance Service.

Rhoades, B. L., Greenberg, M. T., Lanza, S. T., & Blair, C. (2011). Demographic and familial predictors of early executive function development: Contribution of a person-centered perspective. *Journal of Experimental Child Psychology, 108*, 638-662.

Riccio, C. A., Sullivan, J. R., & Cohen, M. J. (2010). *Neuropsychological assessment and intervention for childhood and adolescent disorders.* Hoboken, NJ: John Wiley & Sons, Inc.

Ritter, B., Nelle, M., Perrig, W., Steinlin, M., & Everts, R. (2013). Executive functions of children born very preterm – deficit or delay? *European Journal of Pediatrics, 172*, 473-483.

Robertson, I. H., Ward, A., Ridgeway, V., & Nimmo-Smith, I. (1994). *Test of Everyday Attention.* Bury St Edmunds, UK: Thames Valley Test Company.

Rohde, T. E., & Thompson, L. A. (2007). Predicting academic achievement with cognitive ability. *Intelligence, 35*, 83-92.

Rosenthal., M., Wallace, G., Lawson, R., Dixon, E., Yerys, B. E., & Kenworthy, L. (2013). Impairments in real-world executive function increase from childhood to adolescence in autism spectrum disorders. *Neuropsychology, 27*, 13-18.

Russell, J. (1997). *Autism as an executive disorder.* New York: Oxford University Press.

Rutledge, K. J., van den Bos, W., McClure, S. M., & Schweitzer, J. B. (2012). Training cognition in ADHD: Current findings, borrowed concepts, and future directions. *Neurotherapeutics, 9*, 542-558.

Schafer, J. L. (1999). Multiple imputation: A primer. *Statistical Methods in Medical Research, 8,* 3-15.

Schrank, F. A., McGrew, K. S., & Mather, N. (2014). *Woodcock-Johnson IV.* Rolling Meadows, IL: Riverside.

Schultz, D. P., & Schultz, S. E. (2008). *A history of modern psychology* (8th ed.). Belmont, CA: Wadsworth.

Schumacker, R. E., & Lomax, R. G. (2010). *A beginner's guide to structural equation modeling* (3rd ed.). New York: Routledge Taylor & Francis Group.

Seckl, J. R., & Holmes, M. C. (2007). Mechanisms of disease: Glucocorticoids, their placental metabolism and fetal 'programming' of adult pathophysiology. *Nature Clinical Practice. Endocrinology and Metabolism, 3*, 479-488.

Seckl, J. R. (2008). Glucocorticoids, developmental programming and the risk of affective dysfunction. *Progress in Brain Research, 167*, 17-34.

Shallice, T., & Burgess, P. (1991). Deficits in strategy application following frontal lobe damage in man. *Brain, 114*, 727-741.

Shallice, T., & Burgess, P. (1996). The domain of supervisory processes and temporal organization of behavior. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 351,* 1405-1411.

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

Silver, C. (2000). Ecological validity in neuropsychological assessment in childhood traumatic brain injury. *Journal of Head Trauma and Rehabilitation, 15*, 973-988.

Simon, B. (1972). Models of mind and mental illness in ancient Greece: II. The Platonic Model. *Journal of the History of the Behavioral Sciences, 8*, 389-404.

Sisk, C. L., & Zehr, J. L. (2005). Pubertal hormones organize the adolescent brain and behavior. *Frontiers in Neuroendocrinology, 26*, 163-174.

Slick, D. J., Lautzeniser, A., Sherman, E. M., & Eyrl, K. (2006). Frequency of scale elevations and factor structure of the Behavior Rating Inventory of Executive Function (BRIEF) in children and adolscents with intractable epilepsy. *Child Neuropsychology, 12*, 181-190.

Sonuga-Barke, E. J. (2002). Psychological heterogeneity in ADHD: A dual pathway model of behavior and cognition. *Behavioral Brain Research, 130*, 29-36.

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

Spreen, O., & Strauss, E. (1991). *A compendium of neuropsychological tests:*

Administration, norms, and commentary. New York: Oxford University Press.

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

Swanson, H. L. (1999). Reading comprehension and working memory in learning-disabled readers: Is the phonological loop more important than the executive system? *Journal of Experimental Child Psychology, 72*, 1-31.

Swanson, H. L., & Ashbaker, M. H. (2000). Working memory, short-term memory, speech rate, word recognition and reading comprehension in learning disabled readers: Does the executive system have a role? *Intelligence, 28*, 1-30.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using Multivariate Statistics* (5^{th} ed.). New York: Pearson.

Terman, L. M. (1922). A new approach to the study of genius. *Psychological Review, 29*, 310-318.

Testa, R., Bennett, P., & Ponsford, J. (2012). Factor analysis of nineteen executive function tests in healthy adult population. *Archives of Clinical Neuropsychology, 27*, 213-224.

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

Thurstone, L. L. (1947). *Multiple factor analysis.* Chicago: University of Chicago Press.

Toplak, M. E., Bucciarelli, S. M., Jain, U., & Tannock, R. (2009). Executive functions: Performance-based measures and the behavior rating inventory of executive function (BRIEF) in adolescents with attention deficit/hyperactivity disorder (ADHD). *Child Neuropsychology, 15*, 53-71.

Toplak, M. E., West, R. F., & Stanovich, K. E. (2013). Practitioner review: Do performance-based measures and ratings of executive function assess the same construct? *Journal of Child Psychology and Psychiatry, 54*, 131-143.

Trenerry, M. R., Crosson, B., DeBoe, J., & Leber, W. R. (1989). *Stroop Neuropsychological Screening Test*. Odessa, FL: Psychological Assessment Resources.

Tucker, L. R. (1951). A method for synthesis of factor analysis studies. *Personnel Research Section Report No.984*, Washington, DC: Department of the Army.

Vygotsky, L. S. (1962). *Thought and language*. Oxford, UK: Wiley.

Wayman, J. C. (2003). *Multiple imputation for missing data: What is it and how can I use it?* Paper presented at the annual meeting of the American Educational Research Association in Chicago, IL.

Wechsler, D. (1989). *Wechsler Preschool Primary Scale of Intelligence-Revised.* San Antonio: Psychological Corporation.

Wechsler, D. (1991). *Wechsler Inteligence Scale for Children* (3rd ed.). San Antonio: Psychological Corporation.

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children* (4th ed.). San Antonio: Psychological Corporation.

Wechsler, D. (2014). *Weschsler Intelligence Scale for Children* (5th ed.). San Antonio: Psychological Corporation.

Welsh, M. C., & Pennington, B. (1988). Assessing frontal lobe functioning in children: Views from developmental psychology. *Developmental Neuropsychology, 4*, 199-230.

Weyandt, L. L. (2005). Executive function in children, adolescents, and adults with attention deficit hyperactivity disorder: Introduction to the special issue. *Developmental Neuropsychology, 27*, 1-10.

Wiebe, S. A., Espy, K. A., & Charak, D. (2008). Using confirmatory factor analysis to understand executive control in preschool children: I. Latent structure. *Developmental Psychology, 44*, 575-587.

Willoughby, M. T., Wirth, R. J., & Blair, C. B. (2012). Executive function in early childhood: Longitudinal measurement invariance and developmental change. *Psychological Assessment, 24*, 418-431.

Wilson, M. D. (2012). The life and times of Thomas Willis and his impact on contemporary medicine. *Journal of the History of the Neurosciences, 21*, 127-131.

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

Wuensch, K. L. (2016). *Comparing two groups' factor structures: Pearson r and the coefficient of congruence*. Retrieved from http://core.ecu.edu/psyc/wuensch/MV/FA/FactorStructure-TwoGroups.docx

Zelazo, P. D., & Carlson, S. M. (2012). Hot and cool executive function in childhood and adolescence: Development and plasticity. *Child Development Perspectives, 4*, 354-360.

Zelazo, P. D., Muller, U., Frye, D., & Marcovitch, S. (2003). The development of executive function in early childhood. *Monographs of the Society for Research in Child Development, 68*, 11-27.

Zelazo, P. D., Qu, L., & Kesek, A. C. (2010). Hot executive function: Emotion and the development of cognitive control. In S. D. Calkins & M. A. Bell (Eds.), *Child development at the intersection of emotion and cognition* (pp. 97-111). Washington, DC: APA.