Summer 2021

The Development and Validation of an Automated Screener of Attention

Rachel Bridges

The Development and Validation of an Automated Screener of Attention

BY

Rachel Bridges

Bachelor of Arts
Marist College, 2015

Master of Arts
University of South Carolina, 2018

Submitted in Partial Fulfillment of the Requirements

For the Degree of Doctor of Philosophy in

School Psychology

College of Arts and Sciences

University of South Carolina

2021

Accepted by:

Scott L. Decker, Major Professor

Kimberly J. Hills, Committee Member

Samuel D. McQuillin, Committee Member

Christine DiStefano, Committee Member

Tracey L. Weldon, Interim Vice Provost and Dean of the Graduate School

DEDICATION

This work is dedicated to Richard Laliberte, my grandfather and my friend. I would also like to thank my parents, Michele and Jim, and my brother, Mike. Your love, encouragement, and guidance make all things possible.

ACKNOWLEDGEMENTS

I am indebted to my advisor, Scott Decker, for supporting and challenging me. Thank you also to my wonderful supervisors, teachers, and role models, especially Kim Hills and Kate Flory. I thank Michael Eason, whose attention to detail and persistence made this project possible, as well as Christopher Anzalone and Joseph Ferraracci for being excellent lab mates and, importantly, great friends. I am eternally grateful for my insightful and compassionate cohort, as well as the unwavering support from Tara Kenworthy and Jordan Ezell. Finally, thank you to friends near and far, for endless loyalty, appreciation, and laughter.

ABSTRACT

Previous research has demonstrated a strong relationship between symptoms of ADHD and academic underachievement. Interventions specific to academic deficits in children with ADHD are available, which are most effective if implemented before secondary concerns arise. Performance based screening is one method for determining the need for early intervention, yet extant measures of attention have limitations for the purposes of large-scale screening. The current study evaluated the psychometric properties and guiding conceptual model of a novel instrument of executive functioning—the GNG Screen— which measures response inhibition via a go/no-go paradigm. Results from Rasch modeling and exploratory factor analysis provide preliminary psychometric support for dimensionality and reliability and suggest further revisions to future versions of the instrument. Importantly, dimensionality findings from the current study align with previous evidence indicating EFs are difficult to measure in isolation. Replicating analyses using a more targeted sample of participants, as well as eliminating redundant and/or outfitting blocks should improve dimensionality findings. Further, item difficulty gleaned from Rasch analyses generally support the guiding conceptual model; however, examination of differences in difficulty suggests a reduction in length may be sufficient for capturing the same range of difficulty. Suggestions for future test development and the establishment of expectations for performance are discussed, in addition to directions for future research and clinical implications.

Table of Contents

LIST OF TABLES

List of Figures

Chapter 1: Introduction

Part I: Background

Attention-Deficit/Hyperactivity Disorder (ADHD) is a complex neurodevelopmental condition marked by a persistent display of inattentive, hyperactive, and/or impulsive symptoms, which occur more frequently and severely than typically observed in individuals at an equal stage of development (American Psychiatric Association, 2013). This disorder constitutes the most prevalent psychiatric concern among youth, affecting 5-8% of children (Polanczyk et al., 2007; Willcutt, 2012). The symptoms of ADHD significantly compromise the functioning of children across multiple domains, with impairments often beginning in early childhood and remaining unremitted into adulthood (Lahey et al., 2004; Massetti et al., 2008). The negative impact of ADHD across the lifespan underscores the importance of early identification to inform early intervention (Sonuga-Barke & Halperin, 2010).

School-aged children with ADHD are of particular concern due to the difficulties they face regarding academic achievement. Academic underachievement is among the most notable characteristic associated with ADHD (Frazier et al., 2007), and a child with an ADHD diagnosis will cost the U.S. education system approximately $5,000 annually, in contrast to typically developing children who each cost approximately $300 per year (Robb et al., 2011). Children with ADHD are more likely to receive special education services (Biederman et al., 1996) and are at an increased risk of grade retention (Frazier et al., 2007). Symptoms of ADHD are associated with lower grades across all academic subjects (Barry et al., 2002; Kaufmann & Nuerk, 2008) and poor standardized tests scores in reading and mathematics (Carlson & Tamm, 2000). A meta-analysis demonstrated moderate to large differences in academic achievement between children with ADHD and typically developing controls, particularly on reading measures (Frazier et al., 2007). Moreover, the link between ADHD and academic concerns appears specific to ADHD-related symptomology and is not necessarily explained by comorbid problems (i.e., conduct disorders, learning disorders; DuPaul et al., 2004; Frick et al., 1991; Hinshaw, 1992; Rapport, Scanlan, & Denney, 1999).

The most prevalent treatments for the core symptomology of ADHD consist of medication (i.e., stimulants; Castle et al., 2007; Zuvekas & Vitello, 2012) and behavioral interventions (Dupaul et al., 2007), which are designed to reduce off-task and disruptive behavior; however, the effect of pharmaceutical and behavioral intervention methods on academic achievement is less researched (DuPaul & Eckert, 1997; Raggi & Chronis, 2006). While educational staff may be more aware of externalizing behaviors (i.e., due to their disruptive nature; versus inattentive behaviors), these are not the symptoms that should be targeted when aiming to improve academic performance (DuPaul et al., 2004). Rather, academic interventions for children with ADHD should target symptoms of inattention and underlying cognitive deficits (Pfiffner & DuPaul, 2015). Moreover, interventions to address academic weaknesses in children with ADHD are most effective when provided early (i.e., before the age of 7 years; Sonuga-Barke & Halperin, 2010). In order to inform the implementation of early interventions, a precise method of early identification is needed.

Early Identification & Computer Adaptive Testing

Early identification is important in that it allows the potential prevention of academic and behavioral problems prior to the onset of more severe impairments (DuPaul & Kern, 2011; Sonuga-Barke & Halperin, 2010). Early identification is also critical because interventions for ADHD yield optimal outcomes when implemented early, as brain plasticity is greater during early development (Dawson, 2008). Additionally, interventions are more effective if provided before the underlying disorder is complicated by secondary problems (i.e., social, behavioral, academic). Early identification offers the alternative treatment strategy of implementing a prevention-based approach, with the aim of decreasing both the emergence and persistence of ADHD symptomology (Sonuga-Barke & Halperin, 2010). Although preventative and early intervention methodologies are less common for ADHD (as compared to other disorders [i.e., Autism Spectrum Disorder (ASD)]), recent advances in understanding the developmental trajectory of ADHD may better inform these approaches. Lastly, research has suggested increased success of early interventions when designed to specifically target ADHD symptomology—in contrast to general intervention approaches—and early identification can allow for the appropriate and accurate selection of such intervention methodology (DuPaul & Kern, 2011).

Inattention, impulsivity, and/or high levels of activity (i.e., the core symptoms of ADHD) are relatively typical among most young children, particularly under certain conditions (e.g., unstructured activities; when fatigued). Additionally, early behavioral indicators of ADHD can be indicative of distinct syndromes, such as ASD and anxiety disorders (Spencer, 2006). Thus, when a concern regarding symptoms of ADHD is raised, it may be difficult to ascertain the source of the presenting problem. Screening for psychological concerns is one solution to this assessment issue (DuPaul & Kern, 2011). Screening identifies the need for the administration of additional assessment measures, which may require more resources than the initial screen (e.g., increased time, money, personnel, etc.). For the purposes of screening, it is suggested that liberal thresholds (i.e., 90th percentile) are used for determining the need for further assessment, and thus the presence of the disorder cannot be determined from positive screening results alone (Sonuga-Barke & Halperin, 2010). As such, screening should not replace best practice diagnostic procedures (for a review of standard assessment methods, see Marsh & Barkley, 2009). Regardless, methods of screening can be conducted at a large scale (i.e., universal screening), expanding the reach of more traditional assessment techniques and subsequently informing the implementation of prevention and early intervention services.

Despite the importance of early intervention, measures of ADHD are not routinely administered as part of early screening models (Simmons et al., 2008). Indeed, universal academic screening is often implemented in schools (i.e., through Multi-Tiered Systems of Support [MTSS]); however, screening measures of attention are typically excluded from these models. Currently, screening methods for ADHD often include brief questionnaires completed by caregivers and/or teachers who report on the presence, frequency, and severity of ADHD related behaviors (see DuPaul & Stoner, 2014). These rating methods are constrained by limitations, including the often energetic and/or inattentive behavioral profiles of most young children (including typically developing youth), the inherent biases associated with observational reports, and the resources required to complete ratings at a large scale (i.e., teachers may need to rate multiple students; Sonuga-Barke & Halperin, 2010)). An alternative to behavioral rating measures is the use of performance-based measures, which would eliminate inherent biases and could be sensitive to underlying deficits that may not manifest through day-to-day behavioral presentations. Yet, extant performance-based measures are also fraught with disadvantages, as most valid performance-based measures of attention are (1) individually administered, (2) time consuming, and (3) typically designed for/used as part of diagnostic evaluations. Additionally, schools have finite resources (e.g., staff, time) and often lack sufficient trained personnel for administering performance-based measures at a large-scale.

A significant emerging theme in psychological assessment is the shift of psychological measures from "pen-and-paper" to computerized administration (Maqableh 2015; Naglieri et al., 2004). Per the National Center for Education Statistics (2010), roughly 95 percent of classrooms include computers and most schools have at least one computer lab. Given the adequate technological infrastructure within schools, a performance-based screener for ADHD that is administered via an automated, computerized format would offer a viable alternative assessment method to address the barriers of implementing large-scale early screening. Additionally, an automated format provides standardized administration practices, thereby improving fidelity and reducing the need for qualified examiners to individually assess each student.

Predictors of ADHD & Academic Impairment

In order to achieve the goals of early identification through computerized screening, research must establish reliable performance-based predictors of ADHD symptomology and related academic impairment that can be measured using an automated instrument. Since ADHD is highly comorbid with learning problems, and 20 – 30% of children with ADHD have an associated learning disorder (LD; Biederman et al. 1991; Pliszka 1998), it has been proposed that academic underachievement associated with ADHD is a product of factors indirectly related to the disorder. However, this theory has been contended by research exploring the underlying neuropsychological deficits in children with ADHD, LD, and comorbid ADHD/LD. Korkman and Pesonen (1994) revealed deficits in the control and inhibition of impulses in children with ADHD; deficits in phonological awareness and verbal memory in children with LD; and deficits across all domains in children with ADHD/LD. Moreover, it has been demonstrated that children with “pure ADHD” (i.e., those who do not have a comorbid LD) also experience academic problems, thereby suggesting academic underachievement is not exclusive to children with LD.

Another preliminary explanation for the association between ADHD and academic achievement was variance in IQ (McGee et al., 1992; Sonuga-Barke et al., 1994), as there is evidence for a negative association between symptoms of ADHD and IQ (i.e., higher symptoms of ADHD associated with lower IQ) and research has found that IQ is predictive of academic achievement (Watkins et al., 2007). However, studies that have controlled for intelligence (Diamantopoulou et al., 2007; Barry et al., 2002) indicate children with ADHD demonstrate academic deficits beyond what IQ predicts. These findings are consistent with work that suggests children with comorbid ADHD/intellectual disability (ID) show a reduction in anticipated level of academic achievement—as predicted by IQ—when compared to children with only ID (Simonoff et al., 2007). Thus, although children with ADHD score lower than controls on measures of IQ, this finding neglects to explain impaired academic performance in children with ADHD.

An alternative theory affirms that underachievement observed in individuals with ADHD is driven by cognitive impairments commonly associated with the disorder. Per the DSM-5 (APA, 2013), a diagnosis of ADHD focuses on the behavioral symptoms of inattention, hyperactivity, and impulsivity; however, these criteria have been criticized in that they do not account for the widespread impairment in executive functioning (EF) experienced by individuals with ADHD. There is a considerable body of literature suggesting children with ADHD/poor EF perform worse on measures of academic achievement than do children with ADHD/age-appropriate EF. Biederman and colleagues (2004) found significantly more youth with ADHD had deficits in EF, as compared to typically developing youth. Moreover, youth with ADHD and cooccurring EF deficits were at increased risk for significant impairments in academic achievement. In this investigation, cooccurring symptoms of ADHD and executive deficits were associated with heightened risk for grade retention and decreased academic functioning, as compared to (1) ADHD alone, (2) LD, and (3) IQ. Moreover, Biederman and colleagues concluded these findings provide evidence for early screening of EF. Thorell (2007) suggested preschool youth with ADHD and EF deficits were more likely to experience later academic difficulties. Further, Diamantopoulou and colleagues (2007) demonstrated that ADHD symptoms and poor EF individually predict academic underachievement; however, an interaction was observed by ADHD subtype: high levels of inattention with executive deficits predicted increased special educational need. This is consistent with the wealth of literature highlighting the relationship between symptoms of inattention (but not hyperactivity/impulsivity) and poor academic performance.

Yet, research has suggested executive deficits are not present in all children with ADHD (Sonuga-Barke, 2002). To address this discrepancy, Sonuga-Barke (2002) proposed the Dual-Pathway Model of ADHD, postulating two distinct pathways underlie the behavioral expression of ADHD. One pathway applies to children whose ADHD symptomology is the functional expression of impaired EF. Conversely, the second pathway applies to children whose ADHD symptomology is the manifestation of a motivational style, characterized by "aversion to delay," wherein the child demonstrates critical differences in reward mechanisms. Thorell (2007) explored the relationship between the pathways of this model and early academic skills in kindergarten students. It was demonstrated that delay aversion was related to symptoms of hyperactivity/impulsivity, while weak EF was related to symptoms of inattention. Additionally, it was found that symptoms of inattention were related to academic weaknesses, while symptoms of hyperactivity/impulsivity were not. Moreover, delay aversion was unrelated to early academic skills, yet significant correlations were found between executive deficits and academic skills. Lastly, EF was found to mediate the relationship between inattention and early academic skills.

Altogether, this field of work suggests EF deficits may serve as potential markers for the early identification of ADHD/related academic impairment and that screening children for executive deficits may prevent academic failure (Biederman et al., 2004). Moreover, symptoms of inattention (versus hyperactivity/impulsivity) have been identified as being related to academic underachievement, and EF often mediates the relationship between inattention and academic underachievement (Daley & Birchwood, 2010). In summary, educational screening systems largely employ tests of academic achievement but not measures specific to attention. Screening for early attentional deficits is critical to inform the early, targeted treatment of ADHD. However, current measures of attention contain practical and theoretical limitations for the early detection of ADHD and related impairment. As a solution to this critical assessment concern, the current study offers an automated, computerized instrument of EF. A literature review of EF and common-performance based correlates of ADHD is offered in the following section, which was used to inform the development of the instrument.

Part II: Literature Review- ADHD and Executive Functions

Considerable research has substantiated that specific executive deficits are associated with ADHD (see Boonstra et al., 2005; Antshel et al., 2014). Executive functions are defined as a group of general regulatory processes that guide an individual's thoughts and behaviors (Miyake & Friedman, 2012). Executive processes are many, but essential features include foresight, goal setting, action initiation, self-regulation, cognitive flexibility, attentional control, and working memory. Executive functions develop throughout childhood and adolescence and are associated with functional outcomes in cognitive, emotional, behavioral, and social domains. Many experts agree EF encompasses multiple related sub-functions, given (1) global executive impairment is rare; (2) specific executive processes are associated with distinct neural systems; and (3) distinct EF processes demonstrate differing developmental timelines (Anderson, 2002).

The first five years of life mark a critical period in the development of EF. During infancy and early childhood, essential elements of EF emerge and create an integral foundation for the maturation of higher-order cognitive processes in adolescence and adulthood. Understanding the process in which EF develops is imperative when exploring executive deficits in early childhood in order to inform accurate, age-appropriate measurement. Developmental models of EF have been constructed largely from factor analytic studies using outcome parameters from EF measures (e.g., Levin et al., 1991; Welsh, Pennington, & Groisser, 1991). Results across studies suggest developmental measures of EF generally load onto three to four factors, which become apparent at different stages of development (Diamond, 2013). For an overview of developmental models of EF, refer to Table A.1.

Despite varying terminology, the following terms will employed be in the current study to describe the four core functions commonly noted within this literature: *working memory* (defined as: holding information in mind and mentally manipulating it), *inhibition* (defined as: "being able to control one's attention, behavior, thoughts, and/or emotions to override a strong internal predisposition or external lure;" Diamond, 2013, p. 137), *cognitive flexibility* (defined as: "changing perspectives or approaches to a problem, flexibly adjusting to new demands, rules, or priorities;" Diamond, 2013, p. 137), and *information processing* (defined as: "fluency, efficiency and speed of output;" Anderson, 2002, p. 74). Here, planning and reasoning are considered higher-order EFs, in accordance with previous research (e.g., Collins & Koechlin, 2012). Thus, given the current study's emphasis on early to middle childhood, an examination of higher-order EFs is outside the scope of this paper.

Barkley (1997) argued that impairments accompanying ADHD (i.e., inattention, hyperactivity, and impulsivity) are secondary symptoms that occur as a result of a primary deficit in EF. This theory was established in part from lesion studies that yield behavioral symptoms similar to ADHD when damage occurs in the prefrontal cortex (Stuss et al., 1986; Stuss et al., 2000). In terms of specific executive deficits, the strongest and most homogeneous support has been established for the relationship between ADHD and lower order, “core” EFs (inhibition, working memory, and cognitive flexibility; e.g., Barkley, 2006; Schoemaker, Mulder, Deković, & Matthys, 2013). Executive deficits have been observed in young children before the onset of behavioral symptoms (e.g., hyperactivity, inattention; Sjöwall et al., 2017), and this field of research suggests deficits in EF may serve as potential markers for the early identification of ADHD. Despite growing consensus that ADHD is characterized by deficits across core EF, the literature demonstrates heterogeneity in the specific EF domains associated with the disorder, as well as their strength of association. A literature review is offered to explore variance in ADHD symptoms explained by early EF and provide clarification regarding the utility of EF deficits in predicting the onset of behavioral symptomology associated with ADHD.

Inhibition

As previously defined, inhibition is the ability “to control one’s attention, behavior, thoughts, and/or emotions to override a strong internal predisposition or external lure” (Diamond, 2013, p. 137). A salient stimulus will automatically capture an individual’s awareness, which is referred to as bottom-up attention and is generated through properties of the stimulus (Posner & DiGirolamo, 1998, Theeuwes, 1991); however, an individual can willingly decide to disregard (i.e., inhibit) a specific stimulus and allocate attentional resources to other stimuli given the overall goal. Inhibition is thought to emerge around the age of 4 years, and steadily develop through the age of 11 years, with the most rapid growth seen between the ages of 5 and 7 years(Barkley, 2012). Moreover, by the age of 16 years, inhibitory control is relatively stable and largely equivalent to that seen in adulthood. Per Barkley (1997), inhibition encompasses three interrelated components: the deliberate restricting of a dominant, automatic, or prepotent response (commonly referred to as “action restraint” or “response inhibition;” for clarity, henceforth the current study will use the term response inhibition; Ikeda, Hirata, Okuzumi, & Kokubun, 2010); the stopping of an ongoing response (identified as “action cancellation;” Eagle et al., 2008); and the suppressing of a competing response in order to carry out a primary response (termed “interference control”).

Response inhibition—a primary symptom of ADHD—is commonly measured by go/no-go tasks (Cragg & Nation, 2008), which require an individual to provide a response when a target stimulus is presented and withhold a response when an alternative stimulus is presented. Action cancellation is best captured with measures such as the stop-signal task (Verbruggen & Logan, 2008), in which a “go” stimuli is present across all trials; however, on a minority of trials a “stop signal” is presented concurrently or shortly after the “go” stimuli, prompting the individual to withhold responding. Typical measures of interference control include the Stroop task (MacLeod, 1991), Simon task (Hommel, 2011), and Flanker task (Mullane et al., 2009). On the Stroop task, a “mismatched” stimuli is presented (classically, a color written in text of a differing color). A participant must name one aspect of the stimuli, while inhibiting the competing aspect. On a Simon task, an individual is prompted for a cue press on the left side of a keyboard and for another cue press on the right side of a keyboard; cues are presented at various locations on a computer screen. One cue is presented at a time, and individuals respond less efficiently when the cue is presented on the side of the computer screen opposite to the response side (Hommel, 2011, Lu & Proctor, 1995). The Flanker task requires an individual to attend to a centrally presented cue and ignore the surrounding irrelevant stimuli. When the irrelevant stimuli are mapped to the opposite directional response from the center stimulus (incompatible trials), subjects respond more slowly because of the need to exercise top-down control (Eriksen & Eriksen, 1974).

Meta-analyses including studies of school-aged children and adolescents have revealed that ADHD is associated with response inhibition, in that individuals with more symptoms of ADHD demonstrate greater weaknesses on measures of inhibitory control (with medium to large effect sizes; Nigg, 2005; Pauli-Pott & Becker, 2011; Stefanatos & Baron, 2007; Willcutt et al., 2005). More recent research has revealed tasks of inhibition better predict ADHD with comorbidities than "pure" ADHD (Pauli-Pott et al., 2014). However, Breaux and colleagues (2016) found measures of inhibition, in conjunction with indices provided by a continuous performance task (CPT; Conners, 2001), significantly predict risk for developing ADHD. Relatedly, when exploring inhibition longitudinally, Rennie and colleagues (2014) reported that children with ADHD differed significantly from children without ADHD on measures of inhibition at baseline assessment (age 7 years); however, no significant differences were indicated at two-year follow-up.

Go/No-Go Tasks. Regarding the utility of specific inhibition tasks, several recent studies have employed go/no-go measures of inhibition, with generally consistent results in terms of predicting symptoms of inattention. Moreover, findings from multiple studies reveal performance on go/no-go measures is significantly related to symptoms of inattention, both concurrently (Sjöwall et al., 2015; Brocki et al., 2010) and longitudinally (Sjöwall, 2017). In terms of hyperactivity/impulsivity, Sjowall and colleagues (2015) reported no significant results regarding the effect of hyperactivity/impulsivity on go/no-go performance concurrently, but an effect was indicated longitudinally.

Given the current focus on go/no-go tasks, additional patterns of performance across gender are reported here. Several studies have found no differences in performance on go/no-go tasks when comparing healthy males and females (i.e., without ADHD; Erickson et al., 2005; Li, Zhang, Duann, Yan, Sinha, & Mazure, 2009; Thakkar et al., 2014). Despite these results, a recent study examining go/no-go performance in healthy adults found minor sex differences, in that females generally outperformed males (Sjoberg & Cole, 2018); this is further supported by sex differences in brain activation in areas associated with inhibitory control during go/no-go administration (Roberts et al., 2008). The majority of research in this domain, however, has examined sex difference exclusively in adults with ADHD. For instance, a recent meta-analysis found that sex did not significantly moderate response inhibition—as measured by go/no-go performance—in individuals with ADHD (Wright et al., 2014). There is limited research that utilizes "pure" go/no-go tasks to examine sex differences in children; however, a large body of literature has examined such differences in youth via CPT. A meta-analysis demonstrated that boys are significantly more impulsive than girls (i.e., boys made more commission errors), but no difference with inattention was found (i.e., relatively equal omission errors; Hasson & Fine, 2012). Within-gender analyses indicated that the difference among boys with and without ADHD was significantly larger than the difference among girls with and without ADHD.

Stroop Tasks. Results from Stroop-like tasks have been largely inconsistent. Through a meta-analysis of Stroop performance, van Mourik and colleagues (2005) found small effect sizes, suggesting this task is a weak measure of the underlying neuropsychological deficits of ADHD. Two recent studies indicated an association between Stroop-like tasks and symptoms of inattention (but not hyperactivity/impulsivity; Brocki et al., 2010; Miranda et al., 2015), while one recent study discovered a relationship between Stroop-like tasks and symptoms hyperactivity (but not inattention; Miranda et al., 2015). However, another recent study found that poor performance on Stroop-like measures was related to having a diagnosis of ADHD with comorbid symptoms (Pauli-Pott et al., 2014).

Other Measures of Inhibition. Several current studies have employed the Statue subtest— a measure of inhibition contained in a larger validated battery of executive functioning (The NEPSY; Brooks et al., 2009)— to explore the relationship between inhibitory control and ADHD symptomology. These studies have suggested significant associations between inhibition and behavioral symptomology, and a pattern related to hyperactivity/impulsivity is apparent. Specifically, performance on Statue was significantly related to overall ADHD symptom levels, both individually (Skogan et al., 2013; Zhang et al., 2018; Jacobson et al., 2017) and in combination with other EF measures (longitudinally; Breaux et al., 2016). Jacobson and colleagues (2017) reported Statue performance as the most predictive measure of ADHD status (categorically; when compared to other measures of inhibition). Lastly, Statue was a significant predictor of teacher rated symptoms of hyperactivity/impulsivity (Lavigne et al., 2015).

Working Memory

Working memory (WM) refers to “a limited capacity system allowing the temporary storage and manipulation of information necessary for such complex tasks as comprehension, learning, and reasoning” (Baddeley, 2000, p. 418). Confirmatory factor analytic studies have found support for two distinct WM domains: (1) verbal/numerical and (2) figural/spatial; different WM tasks measure distinct domains of WM content. Verbal/numerical working memory is often measured by complex span tasks, such as backward-digit span, and tasks that require the reordering of verbal stimuli (Barrouillet et al., 2009; Conway et al., 2005; Daneman & Carpenter 1980). Additionally, n-back tasks are often used as a measure of verbal/numerical working memory, in which participants are asked to indicate if stimuli provided in a string coordinate with previous items. Similarly, figural/spatial working memory is often measured through tasks that require the recalling and/or reordering of non-verbal stimuli. For example, a common measure of figural/spatial WM is the Corsi Block test (Lezak 1983), in which an individual must tap a series of the blocks in the same order as the examiner. Computerized versions of this task are included in the Automated Working Memory Assessment battery (AWMA; Alloway, 2007; Alloway et al., 2009) and the CANTAB (Luciana & Nelson, 2002).

Nigg and colleagues (2012) reported the highest effect sizes for working memory in a recent meta-analysis of ADHD related symptoms; however, Pauli-Pott and Becker (2011) noted small mean effect sizes for working memory in predicting ADHD. Several recent studies have established a relationship between working memory ability and

ADHD group membership status (Sjowell & Thorell, 2019; Gremillion et al., 2018; Rennie et al., 2014), as well as dimensional symptoms of inattention and hyperactivity/impulsivity (Skogan et al., 2013). It is noted, however, that Rennie and colleagues (2014) reported an effect only at two-year follow-up, but not at baseline assessment (age 7 years). Conversely, Zhang and colleagues (2018) found working memory did not predict ADHD group membership; however, this study only included children between the ages of 4 and 5 years.

Verbal/ Numerical Working Memory. In terms of distinct working memory types, multiple recent studies have found verbal/numerical working memory to significantly differentiate individuals with ADHD from typical controls, as well as to contribute significant variance in predicting overall ADHD symptoms (Gremillion et al., 2018; Sjowall & Thorell, 2019; Skogan et al., 2013). Additionally, verbal working memory has been reported to predict symptoms of inattention (Miranda et al., 2015; Brocki et al., 2009) and symptoms of hyperactivity (Miranda et al., 2015). However, one study found verbal/numerical working memory to be insufficient in differentiating between subtypes of ADHD (Zhang et al., 2018), and another indicated verbal working memory is unrelated to hyperactivity/impulsivity (Brocki et al., 2009).

Figural/Spatial Working Memory. Regarding figural/spatial working memory, less consistent support has been suggested in recent studies. Results have yielded minimal support for the relationship between non-verbal measures of working memory and symptoms of inattention both concurrently (Miranda et al., 2015) and longitudinally (Sjowall et al., 2015). Yet, Breaux and colleagues (2016) found measures of non-verbal working memory approached significance in differentiating children with pure ADHD from children with ADHD/comorbidity.

Cognitive Flexibility

Cognitive flexibility is defined as “the ability to shift between response sets, learn from mistakes, devise alternative strategies, divide attention, and process multiple sources of information concurrently” (Anderson, 2002, p. 74). This core EF is built on inhibition and working memory and is often attainted later in development than the aforementioned executive processes. Cognitive flexibility is typically measured using a variety of task-switching and set-shifting measures, and a classic task in this domain the Wisconsin Card Sorting Task (WCST; Milner, 1964; Stuss et al., 2000). Each card in this task can be organized by multiple domains (i.e., color, shape, or number). The goal is for the participant to determine the accurate organizational condition based on provided feedback and to adaptably change organizational criteria based on this feedback. Zelazo and colleagues created a simple measure of task switching called the Dimensional Change Card Sort (DCCS). During this task, an individual is instructed to sort six cards according to one dimension (e.g., color) and subsequently sort the cards according to a second dimension (e.g., shape). This task intentionally minimizes memory demands through visual cue and verbal reminders from the examiner. Another task that measures this domain is the Trail Making Test, which is a timed task that prompts an individual to connect a string of letters and numbers in order while switching between numbers and letters. Lastly, an alternative group of tasks that measure cognitive flexibility comprises design fluency, verbal fluency, and semantic fluency. In these tasks, the most typical answer often comes to mind first; however, individuals with more cognitive flexibility can provide alternative, creative answers (Diamond).

In a meta-analysis, Willcutt and colleagues (2005) found a weaker relationship between ADHD status and perseverative errors, as measured by the WCST (as compared to other domains of EF). Additionally, the majority of studies included in this meta-analysis did not detect group differences on the Trail Making test. More recent studies examining the relationship between ADHD and EF have not found cognitive flexibility significant in predicting symptoms of ADHD when these tasks have been explored in isolation (Montamedi et al., 2015; Sasser et al., 2014)

Information Processing

Per Anderson (2002), information processing is defined as the fluency, efficiency, and speed of output. Measures of reaction time (e.g., reaction time variability [RTV]) are often gathered through continuous tasks associated with other domains of EF, such as go/no-go tasks (Anderson, 2000). While studies often do not employ Anderson's term "information processing" to describe this domain of EF, it apparent when this domain is being measured based on the metrics employed within studies. Elevated RTV among children with ADHD, versus control children, has been demonstrated across several studies using a variety of computerized tasks (Tamm et al., 2012). Between-group differences on RTV tend to be larger in magnitude than other neuropsychological indicators (e.g., delay aversion tasks; Epstein et al., 2011). Two previous studies have examined the relationship between RTV and ADHD, independent from other domains of EF. One study found independent effects of working memory, RTV, and delay aversion

(Kuntsi, Oosterlaan, & Stevenson, 2001), while another found independent effects of RTV and inhibition only (Wåhlstedt, Thorell, & Bohlin, 2009).

Two recent studies reported individuals with ADHD demonstrate increased RTV on computerized tasks as compared to individuals without ADHD (Sjowall et al., 2019; Cak et al., 2017); however, when individual symptom categories were explored, results were largely inconsistent. Cak and colleagues (2017) suggested correlations between measures of RTV and symptoms of hyperactivity, but not symptoms of inattention. These results are supported by Barnard and colleagues (2018), who found reaction time related to symptoms of hyperactivity in males. In contrast, Rezazedah and colleagues (2011), who employed a sample of all males, did not find effects involving RTV. Additionally, two recent studies found RTV significantly related to inattention both concurrently (Sjowall et al. 2015) and longitudinally (Sjowall et al., 2017), but no effect on symptoms of hyperactivity/impulsivity.

Limited meaningful findings regarding information processing have been reported beyond RTV. Yet, two recent studies suggested response speed is related to hyperactivity (Barnard et al., 2018; Rezazedah et al., 2011). Interestingly, unusually fast reaction time on continuous tasks was found related to increased externalizing problems—which subsume hyperactivity—in girls only (Barnard et al., 2018). Additionally, speed on a visual search measure was significantly related to symptoms of hyperactivity, but not cognitive problems/ inattention (Rezazadeh et al., 2011)

Implications for Screening

While evidence for between-group differences on measures of EF has been demonstrated across all four core executive domains, findings have been largely variable by task type. Although some evidence has been demonstrated for tasks of working memory, this domain has also been demonstrated as a core deficit in children with LD (Korkman & Pesonen, 1994). Thus, working memory screeners may not have sufficient sensitivity to identify children with symptoms of ADHD and related academic underachievement. Support has also been found for tasks of inhibition, yet only specific task types have yielded strong and consistent associations with ADHD symptomology. Go/no-go measures of response inhibition have been strongly related to symptoms of inattention and consistently predicted group membership for individuals with primarily inattentive presentations. Evidence for the association between go/no-go measures and inattentive symptoms provides support for the employment of response inhibition measures when screening young-children for symptoms of ADHD. Moreover, this association does not appear to differ between computerized and non-computerized measures of inhibition, providing additional support for use of an automated instrument.

Currently, there are no computerized measures of response inhibition designed or suitable for use within school-based, early screening systems. Screeners are designed to be brief, as well as easily administered and scored; thus, certain tasks may be difficult to incorporate into traditional screening procedures. Similar to extant measures of behavioral ratings, current performance-based measures of EF are not designed for screening purposes. Moreover, if extant tasks were administrated individually—in addition to larger screening batteries that assess for academic concerns—it would be improbable to expect “screening” to occur at a large scale. While computerized measures of executive functioning exist (i.e., the Tests of Variable Attention [TOVA; Forbes et al., 1998]; The Conners CPT [Conners, 2001]; The Integrated Visual and Auditory CPT

[IVA; Tinius, 2003]; the Auditory CPT [ACPT; Riccio, 1996]; and the Gordon Diagnostic System [Dickerson et al., 2001]), as well as computerized batteries that contain measures of executive functioning (the CANTAB [DeLuca et al., 2003]; the MicroCOG [Elwood, 2001]; and CNS Vital Signs [Gualtieri & Johnson, 2003]), these instruments are either not fully automated, not designed as screening measures, and/or do not include theoretically based measures of response inhibition. Thus, the present study focuses on the development of a new instrument that aims to address the major limitations of currently available instruments.

Part III: Current Study

To date, no study has examined a fully automated screener of response inhibition specifically for use of the early identification of ADHD symptomology. Thus, the current project aims to address this critical assessment need by developing, scaling, and validating a go/no-go screener, henceforth referred to as the GNG Screen, to be used in conjunction with academic screeners within MTSS. Given the developmental trajectory of EF, as well as the typical timing of early identification and intervention, the target age range for GNG Screen is ages 4 through 8 years.

Guiding Conceptual Model

While children with ADHD are assumed to show a deficiency regarding all three interrelated aspects of inhibition (i.e., response inhibition, action cancellation, and interference control), the current review of literature suggests performance on go/no-go tasks—which measure response inhibition (not action cancellation or interference control) —is correlated with and/or predictive of ADHD symptomology. Moreover, performance on these tasks is highly related to symptoms of inattention, which are more strongly correlated with academic underachievement than symptoms of hyperactivity/impulsivity. The traditional go/no-go paradigm involves two stimuli: a “go” stimulus and a “no-go” stimulus. Participants are prompted to respond quickly—typically through a button-press—to the presentation of “go” stimuli exclusively, and response inhibition is measured by appropriately withholding responds to “no-go” stimuli. Number of commission errors (i.e., responding to a “no-go stimulus”) is traditionally used as a measure of poor inhibitory control (Sjowall, 2015).

For exploratory purposes, omission scores, percentage of correct trials, response time, response time variability, and efficiency— which measure aspects of sustained attention and information processing—were also calculated. Although these metrics are thought to measure distinct EFs (i.e., not response inhibition), individual functions tend to vary at similar levels *within* people (Barkley, 2012). For example, an individual with relatively higher sustained attention will also likely have relatively higher levels of inhibitory control and greater capacity for information processing. Thoroughly examining performance on these metrics is outside the score of the current study; however, this should be considered in future research, as discussed in Chapter 4.

Building a Prepotent Response: “Go” to “No-Go” Ratio. A fundamental element of a go/no-go task is the weighting towards go stimuli (i.e., more “go” stimuli” than “no-go” stimuli), thereby building the prepotent inclination to respond, consequently increasing the inhibitory effort necessary to successfully withhold responding to “no-go” stimuli. This component of a go/no-go task is distinct from a continuous performance task—a similar measure of inhibition that prompts the individual to produce a rapid response when cued and later inhibit a response when an alternative cue is presented—

which contains a *minority* “go” response. Thus, tasks from a go/no-go paradigm better fit the requirement for measures of response inhibition to build a prepotent response (Barkley, 1997; Berlin & Bohen, 2002). Given this central feature, all blocks within the current measure of inhibition were weighted towards “go” stimuli. Berlin and Bohlin’s (2002) go-rate of 70%, which was applied within a go/no-go task created for children, was used in the current study.

Inter-stimulus Interval. In addition to ratio of “go” to “no-go” items, inter-stimulus interval has been shown to strengthen the prepotent tendency to respond and thereby increase inhibitory effort during “no-go” trials (Votruba & Langenecker, 2013). In other words, a stronger prepotent response to “go” cues is built when stimuli are presented in closer proximity. Relatedly, allocating attention to an initial cue momentarily deprives an individual the attention needed for a subsequent cue (termed the “Attentional Blink;” Ashcraft & Radvansky, 2014). Studies have used this paradigm to explore attentional deficits in individuals with ADHD (Hollingsworth et al., 2001; Li et al., 2004). When compared to healthy controls, individuals with attention deficits demonstrate a wider attentional blink. Moreover, individuals with poor attention demonstrate worse performance on go/no-go tasks, and more broadly on measures of rapid responding in which stimuli are presented in close proximity. Lastly, go/no-go tasks designed for older youth, adolescents, and adults generally contain a shorter inter-stimulus interval than tasks designed for younger children (Berlin & Bohlin, 2002; Votruba & Langenecker, 2013). Given the developmental progression of EF, this further suggests tasks with shorter inter-stimulus interval require additional inhibitory effort to accurately respond.

Given the literature concerning this domain, it is proposed inter-stimulus interval contributes to the building of a prepotent response—albeit secondary to the aforementioned “go” to “no-go” ratio—and is ultimately related to the inhibitory demand of a given item. Thus, the current model ascertains items with shorter inter-stimulus interval require more inhibitory resources (versus items with a relatively longer inter-stimulus interval) and are thus considered more difficult. Several go/no-go tasks designed for young children have an inter-stimulus interval of 5000 msec (Berlin & Bohlin; 2002), whereas tasks designed for older children, adolescents, and adults have an inter-stimulus interval ranging from 500 msec to 1000 msec. It is noted, however, inter-stimulus interval tends to vary across tasks, regardless of the targeted level of development (i.e., some tasks designed for adults contain longer inter-stimulus intervals than tasks designed for children). Nonetheless, the current model proposes two levels of inter-stimulus interval which were included in the task: short (500 msec) and long (5000 msec).

Discrimination and Distractors. Beyond the building of a prepotent response, research from cognitive science has suggested two additional factors that may impact the inhibitory effort necessary to successfully respond within the go/no-go paradigm. First, a significant source of intrusive and extraneous information is needed for inhibitory control to adequately function (Ashcraft & Radvansky, 2014). In other words, if irrelevant information is “strong and wrong,” less inhibitory effort is required for withholding a response. A classic demonstration of this phenomena is offered by Tipper (1985), who found individuals were slower to respond to target trials when a trial was preceded by a similar distractor (e.g., same object in different color) than trials where the ignored stimulus was a distinct object. Similarly, research has established that individuals are able to accurately respond to a cued stimulus when distractors are more noticeably different (Ashcraft & Radvansky, 2014), and individuals respond more efficiently to simple stimuli that are more commonly encountered in day-to-day life. Moreover, in a go/no-go task designed for use within research, Berlin and Bohlin (2002) created separate versions for younger children (involving simple shapes) and older children (involving more complex patterns) to account for developmental increases in response inhibition. Moreover, to increase inhibitory effort within both versions, Berlin and Bohlin (2002) added additional distractors.

Together, this evidence suggests two additional factors that increase the inhibitory demand of an item: 1) increased competing stimuli that are similar to the target stimulus (i.e., distractors), and 2) increased design complexity that is more difficult to discriminate. In terms of distractors, most go/no-go paradigms contain two stimuli; however, complex go/no-go paradigms often contain no more than four stimuli. Thus, the current model proposes two levels of distractors that were included in the current task: low (two stimuli) and high (four stimuli). Regarding discrimination, in line with Berlin and Bohlin (2002), two levels of discrimination were included in the current study (low and high).

Blocks. In order to capture each of the aforementioned factors related to the building of a prepotent response, the GNG Screen was organized into eight blocks. Each block captures a distinct combination of the above factors (e.g., *Block 3: High Discrimination, Short Inter-stimulus Interval, and Low Number of Distractors*). Blocks were generally presented in order of theoretical difficulty, from easier to more difficult blocks; thus, more inhibitory effort is expected to provide a correct response on later blocks. Refer to Tables A..2 and A.3 for a graphical organization of block order. Given Berlin and Bohlin (2002) developed separate go/no-go tasks based on level of discrimination to account for developmental differences in inhibition, discrimination will be the highest rank of organization. Thus, blocks will be organized into two levels: *Low Discrimination* and *High Discrimination*. Blocks in *Low Discrimination* are hypothesized to be more difficult than blocks in *High Discrimination*.

The inter-stimulus interval of go/no-go tasks also varies by level of development, albeit less consistently and clearly than discrimination of stimuli. Namely, tasks with shorter inter-stimulus intervals are often, although not always, contained in tasks designed for older individuals. Thus, inter-stimulus interval was the next rank of organization, with each level (*Low Discrimination* and *High Discrimination*) including two blocks that contain items with a relatively long inter-stimulus interval and two blocks with a relatively short inter-stimulus interval. Based on this ranking, blocks in *Low Discrimination* are hypothesized to be more difficult than blocks in *High Discrimination*, regardless of inter-stimulus interval; however, within these levels, blocks with a short inter-stimulus interval are proposed to be more difficult than blocks with a long inter-stimulus interval.

The lowest rank of organization is number of distractors, as tasks designed for both younger and older children include a high number of distractors as a method of increasing inhibitory effort (Berlin & Bohlin, 2002). Thus, within each level (*Low Discrimination* and *High Discrimination*), blocks with a short inter-stimulus interval are proposed to be more difficult regardless of number of distractors; however, a high number of distractors will be more difficult than a low number of distractors. For example, a block with *High Discrimination*, *Short Inter-stimulus Interval*, and *Low Number of Distractors* is considered more difficult than a block with the features *High Discrimination*, *Long Inter-stimulus Interval*, and *High Number of Distractors*.

Broad Objectives and Research Questions

The current study uses theoretical models of response inhibition to develop an automated test of inhibitory control that can be used within universal screening models. The proposed study investigates the validity of theoretical specifications used for constructing the response inhibition test and potential for enhancing early identification of children with ADHD. This research has significant potential to influence early screening practices, including screening systems currently implemented within schools. Validated measurement of attention deficits using the proposed testing procedures—which assesses early executive dysfunctions through the measurement of response inhibition— has tremendous value given the educational implications associated with ADHD.

To achieve the goals of the current study, items were developed in accordance with the guiding conceptual model. Program de-bugging and modifications to test items were conducted and the GNG Screen was administered to a sample for initial validation. Analyses were conducted examining the potential effects of demographic factors, such as age and gender, on test performance. Given the aforementioned trends in the development of EF, the current study includes participants divided into the following age groups, which largely follow key periods of EF growth: Early Childhood (ages 4 - 6 years), Middle Childhood (ages 7 - 8 years), Late Childhood (ages 9 - 11 years), Early Adolescence (ages 12 - 15 years), Late Adolescence (ages 16 - 19 years), and Early

Adulthood (ages 20 + years; Barkley, 2012). Further, the relationship between exploratory metrics of the GNG screen were examined via Pearson's correlation coefficients. Rasch modeling and exploratory factor analysis (EFA) were then applied to evaluate dimensionality and reliability of the items, blocks, and overall instrument. These analyses provided psychometric support for the measure. Previous studies of both traditional (pen-and-paper) and computerized executive functioning, including go-no/go tasks, have also explored item difficulty using Rasch modeling (e.g., Ferreira et al., 2011; Pomplun & Custer, 2005). Thus, Rasch modeling was also used to test predictions from the guiding conceptual model regarding item difficulty, which offer empirical support for the guiding conceptual model. Results from this study, at large, will guide future test development and establish empirical expectations for task performance. Primary research questions are outlined below and discussed further in Chapter 2.

Research Question 1: What is the dimensionality of the GNG Screen?

In the current study, response inhibition was the general dimension hypothesized to underlie the GNG Screen. Moreover, the test as a whole, as well as each block within the subtest, was designed to be unidimensional. To assess the validity of the theoretical model, as well as check the assumptions of the Rasch model to inform interpretation of results from Research Question 2, the following questions were asked:

<u>Research Question 1a. Evaluating Unidimensionality.</u> The Rasch model assumes measures are unidimensional, meaning the instrument only measures a single underlying construct. Evidence of unidimensionality— gleaned from Rasch analyses—provide support for the organization of the GNG Screen. Additionally, results from Rasch

analyses offer guidance for further test development by highlighting items and/or blocks that may need revision due to violations of the unidimensionality assumption.

Research Question 1b. Evaluating Block Factor Structure. In addition to Rasch modeling, EFA was used to further assess the factor structure of the GNG Screen. More specifically, results from factor analysis provide information regarding the latent trait underlying performance across blocks. Further, results provide guidance for future instrument development regarding the inclusion of distinct blocks.

Research Question 2: Do items progress in difficulty according to theoretical expectations?

Research Question 2a. Evaluating Item Difficulty. Based on the guiding conceptual model, it was predicted that blocks with shorter inter-stimulus interval, lower stimulus discrimination, and/or higher number of distractions will contain more difficult items; items within a given block are hypothesized to be of similar difficulty. This is important to examine in order to determine the range of difficulty within the current version of the measure, as well as potential redundancy across levels of difficulty.

Research Question 2b. Evaluating Item Fit. To further inform future test development, item fit indices were examined. These values helped identify items that were not corresponding with the expected pattern of performance, indicating they may need to be revised or omitted. Further, item fit values provided additional psychometric support for the unidimensionality of the instrument.

Research Question 2c. Evaluating Item/Person Reliability. To determine the replicability of items and individuals, and provide additional psychometric support for the GNG Screen, reliability estimates were calculated using the Rasch model.

Research Question 2d. Evaluating Floor/Ceiling Effects. To identify potential items missing in the lower or upper end of the scale, the presence of floor and ceiling effects were examined. Results will inform future test development.

Chapter 2: Method

Participants

The current study included 119 participants between the ages of 4 and 32 years (M = 19; SD = 8.25). While the GNG Screen is designed for use with school-aged children (aged 4 to 8 years), a varied sample of ages and abilities is adequate for the current aims, as increased variability allows for improved calibrations within the Rasch model. Given the developmental nature of EF, as discussed within Chapter 1, participants were categorized within the following age groups: Early Childhood (ages 4 - 6 years), Middle Childhood (ages 7 - 8 years), Late Childhood (ages 9 - 11 years), Early Adolescence (ages 12 - 15 years), Late Adolescence (ages 16 - 19 years), and Early Adulthood (ages 20 + years). Convenience sampling was used and participants for pilot administration were recruited at the University of South Carolina through announcements in undergraduate and graduate level classes, as well as word-of-mouth, and advertisements at public schools, media outlets, and other community locations throughout the Columbia, South Carolina metropolitan area. Undergraduate students received course credit for participation.

Instruments

Inhibition Test (The GNG Screen)

Following construct definition and item development, the GNG Screen was created and administered using Unity software. Unity is a desktop application purposed of the development of computer-based games capable of being deployed across a variety platforms. It, in essence, affords the dynamic presentation of visual and audio stimuli, enabling the user to interact with stimuli. The GNG Screen was tested for compatibility on Windows operating systems prior to data collection. The task was not compatible with Mac or Linux operating systems.

The GNG Screen includes two levels, each with four blocks, largely completed in order of ascending difficulty as defined and hypothesized by the guiding conceptual model. After initial item development blocks were reordered to maintain the unidimensional nature of the task, as switching between different rules for responding increases working memory and/ or cognitive flexibility load. See Tables A.2 and A.3 for the order of hypothesized difficulty, as well as the order of administration within the current version. Henceforth, blocks will be referred to in order of administration. Responses were provided by screen-press on a designated area of a touch screen computer, press of the space bar, or mouse touch (i.e., touch the "go" stimulus).

All information, including task instructions, was presented by a recorded voice or a visual presentation on the computer screen. The program was designed with learning trials for Blocks 1, 3, 5, and 7. As the rules for responding align with the aforementioned blocks, learning trials were not provided for blocks 2, 4, 6, and 8 (i.e., short interstimulus interval)—which were presented consecutively with the corresponding long interstimulus interval block. If a participant did not correctly respond within three learning trials, the task was programmed to discontinue. Following learning trials, the program provided an audio prompt to "press the green button" in order to begin the testing trials. The program was designed to provide a "child-friendly" prompt if a button press did not occur within

15 seconds (i.e., character appeared on the screen asking "Are you still there?"). After three prompts are provided with no response, the task was programmed to discontinue. Stimuli were centered on the screen across trials. A white screen was presented between stimuli. A screen that read "Great Job!" with a smiling character, gesturing a "thumbs-up" was depicted after completion of all blocks. To prevent erroneously exiting prior to task completion, the user was instructed to press the "down arrow" and "e" keys simultaneously to exit the administration; pressing these keys allowed for exiting the task at the test at the end of administration or for discontinuation at any point throughout the administration. The total administration was approximately 18 minutes. Item correctness for both "go" and "no-go" items were pulled to an online database (0 = Incorrect, 1 = Correct). Of note, the blank screen following a given item was included within the item's score. For example, if a participant pressed the button while the blank screen was present directly after a "go" stimulus, the item would be counted as correct even if the participant did not press the button while the stimulus was present. Likewise, if a participant pressed the button while the blank screen was present directly after a "no-go" stimulus, the item would be counted as incorrect, even if the participant did not press the button while the stimulus was present. In addition to item correctness, reaction time was collected for each button press.

Difficulty Parameters. Within the GNG Screen, participants generally started with blocks hypothesized to be easier (i.e., requiring the least demand on inhibitory control) and generally progressed toward hypothetically more difficult blocks. The only exception is Blocks 2 and 3, and Blocks 6 and 7, which were reordered to adhere to unidimensionality assumptions, as noted above. According to the guiding conceptual model, block difficulty was operationalized as inter-stimulus interval, number of distractors, and level of item discrimination.

Blocks. The task contained 8 blocks, which measure each combination of factors associated with block difficulty as defined by the guiding conceptual model (see Tables A.2 & A.3). Each block began with a tutorial screen and unscored trial items. In accordance with Berlin and Bohlin (2002), 30 items were included within each block with a set go-rate of 70%. Items were randomized within blocks and the same sequence of items was presented to each participant.

<u>Level 1. High Discrimination</u>**.** Blocks contained in Level 1 are visually represented in Table A.2.

<u>*Block 1. Long Inter-stimulus Interval and Low Number of Distractors.*</u> Block 1 (difficulty level 1) included two stimuli (blue square; red circle). Examinees were instructed to press the computer screen (“go”) when the blue square was presented but to make no response (“no-go”) when the red triangle was presented. In accordance with Berlin and Bohlin (2002), each stimulus was presented for 800 msec, with an interstimulus interval of 5000 msec.

<u>*Block 2. Short Inter-stimulus Interval and Low Number of Distractors.*</u> Block 2 (difficulty level 3) included two stimuli (blue square; red triangle). Examinees were instructed to press the screen (“go”) when the blue square was presented but to make no response when the red triangle was presented (“no-go”). Each stimulus was presented for 800 msec with an interstimulus interval of 500 msec between shapes.

<u>*Block 3. Long Inter-stimulus Interval and High Number of Distractors.*</u> Block 3 (difficulty level 2) included four stimuli (blue square; red square; blue triangle; red

triangle). Examinees were instructed to press the computer screen ("go") when a square was presented but to make no response ("no-go") when a triangle was presented, irrespective of color. Each stimulus was presented for 800 msec with an interstimulus interval of 5000 msec.

Block 4. Short inter-stimulus interval and High Number of Distractors. Block 4 (difficulty level 4) included four stimuli (blue square; red square; blue triangle; red triangle). Participants were instructed to press the screen ("go") when a square is presented but to make no response when a triangle was presented ("no-go"), irrespective of color. Each stimulus was presented for 800 msec with an interstimulus-interval of 500 msec.

Level 2. Low Discrimination. Blocks contained in Level 2 are visually represented in Table A.3.

Block 5. Long Inter-stimulus Interval and Low Number of Distractors. Block 5 (difficulty level 5) included two stimuli (simple blue square; blue square with horizontal line). Examinees were instructed to press the screen ("go") when the blue square with horizontal line is presented but to make no response when the simple blue square was presented ("no-go"). Each stimulus was presented for 800 msec with an interstimulus interval of 5000 msec.

Block 6. Short Inter-stimulus Interval and Low Number of Distractors. Block 6 (difficulty level 7) included two stimuli (simple blue square; blue square with horizontal line). Examinees were instructed to press the screen ("go") when the blue square with horizontal line was presented but to make no response when the simple blue square was presented (“no-go”). Each stimulus was presented for 800 msec with an interstimulus interval of 5000 msec.

Block 7. Long Inter-stimulus Interval and High Number of Distractors. Block 7 (difficulty level 6) included four stimuli (simple blue square; blue square with horizontal line; simple red square; red square with horizontal line). Examinees were instructed to press the screen (“go”) when a square with horizontal line was presented but to make no response when a simple square was presented (“no-go”), irrespective of color. Each stimulus was presented for 800 msec with an interstimulus interview of 500 msec.

Block 8. Short inter-stimulus interval and High Number of Distractors. Block 8 (difficulty level 8) included four stimuli (simple blue square; blue square with horizontal line; simple red square; red square with horizontal line). Examinees were instructed to press the screen (“go”) when a square with horizontal line was presented but to make no response when a simple square was presented (“no-go”), irrespective of color. Each stimulus was presented for 800 msec with an interstimulus interval of 500 msec.

Scoring. In line with Barkley’s (1997) theory of ADHD, commission errors were considered a direct measure of inhibitory control. Thus, in the current study, commission errors were the primary metric gleaned from performance on the GNG Screen; however, additional exploratory analyses involving other metrics were also conducted (described below). In terms of commission errors, responses to each “no-go” item were scored “1” if correct (i.e., the participant did not respond), and “0” if there was an error of commission. As described previously, each block included 30 items with a go-rate of 70%. As such, commission scores range from 0 to 9 within a given block, and 0 to 72 for the entire measure. Of note, higher scores indicate higher accuracy (i.e., less commission errors).

After raw data were retrieved from the online database, *Total Commission Score* and *Block Commission Score* across all 8 blocks were calculated for each participant using Microsoft Excel.

For exploratory purposes, omission scores, percentage of correct trials, response time, response time variability, and efficiency were also calculated. Person's correlation coefficients were calculated to examine the relationship between metrics.

<u>Omission Scores:</u> Errors of omission (withholding a response on "go" trials) are considered a measure of sustained attention (Berwid et al., 2005). In the present study, responses to each "go" item were scored "1" if correct (i.e., the participant did respond) and scored "0" if there was an error of omission. Of note, higher scores indicate higher accuracy (i.e., less omission errors). After raw data were retrieved from the online database, *Total Omission Score* and *Block Omission Score* across all blocks were calculated for each participant using Microsoft Excel.

<u>Percentage of Correct Trials:</u> In line with Berlin and Bohen (2002), the percentage of correct trials (PCT) was computed by dividing the number of correct responses (across both "go" and "no-go" items) by the number of possible responses within a given block of the task (30 total items), as well as across the entire task (240 total items). *Block PCT* across all blocks and *Total PCT* were calculated for each participant using Microsoft Excel.

<u>Response Time and Response Time Variability:</u> Average response time (RT) was calculated in Microsoft Excel using the mean time for responding, in seconds, for each item within a given block (*Block RT*), as well as across the entire task (*Total RT*). Further, the standard deviation of response time was computed in Microsoft Excel as measure of

Response Time Variability (*RTV*) across all blocks (*Block RTV*), as well as across the entire task (*Total RTV*).

Efficiency: Efficiency is a measure that considers both response time and accuracy, such that participants who respond accurately and rapidly yield the highest efficiency scores. Individuals who respond accurately and slowly, or less accurately and rapidly, have proportionately lower scores (Gur et al., 1992; Langenecker et al., 2005), and individuals who respond less accurately and slowly will yield the lowest efficiency scores. In line with Weidacker and colleagues (2017), the efficiency score for each block (*Block Efficiency*), as well as across the entire task (*Total Efficiency*) was computed in Microsoft Excel using the following formula: PCT/RT.

Demographic Analyses. A two-way analysis of variance (ANOVA) was conducted to determine whether gender, age group, or interactions among these factors significantly affected *Total Commission Score*. Due to sample limitations, ethnicity was excluded from this analysis, which is further described when discussing limitations in Chapter 4. Additional exploratory two-way ANOVAs were conducted to determine whether gender, age group, or interactions among these factors significantly affected *Total Omission Score*, *Total PCT*, *Total RT*, *Total RTV*, and *Total Efficiency*.

Procedure

The GNG Screen was designed with the capability for administration and scoring online, without need for an examiner. In the current study, it was administered in a laboratory setting with supervision by a trained research assistance, as well as in remote conditions. Approval from the University of South Carolina's IRB was received prior to all data collection.

For both laboratory and remote administration, interested participants (and/or guardians, when applicable) were directed to a project website, where they were prompted to watch an instructional video outlining consent/assent procedures, instructions for participation, and navigation of the testing platform. For laboratory administration, undergraduate and graduate level participants completed testing at the University of South Carolina in Columbia, South Carolina. The Applied Cognitive Neuropsychology (ACN) laboratory provided equipment (i.e., touch screen computers) necessary for testing. Participants provided general demographic information to study staff, who created a unique testing account for participants on the project website. Participants used the account to log into the testing interface, which was pre-downloaded onto laboratory computers. For remote administration, participants independently registered for an account on the study website, provided demographic information, downloaded the GNG Screen application, and used registration credentials to access the interface. Once within the interface, the GNG Screen is entirely automated. Test data, along with participant demographic information, and item information were recorded in an online, password-protected database. After test completion, data were screened for excessive response times, erratic responding, and incomplete administration.

Analytic Procedures

Data were extracted from the online, password-protected database into Microsoft Excel for initial data organization, screening, and computation of test metrics. There were no missing data, as all participants completed the entire GNG Screen. Data were imported into Jamovi (The Jamovi Project, 2021), Winsteps(Linacre, 2021), and IBM SPSS Statistics for Macintosh, Version 25 (2020) for remaining analyses. The Rasch model was the primary method of data analysis. Thus, an overview of the Rasch model is offered below, including model assumptions. In addition to the aforementioned exploratory demographic and correlation analyses, primary analytic procedures are detailed below, organized by the research objective.

The Rasch Model

Per DiStefano and Morgan (2010), Rasch modeling is a one-parameter item response theory (IRT) model that employs scores obtained from an instrument to compute the likelihood that an individual will accurately respond to a given item, based on the individual's ability and the item difficulty. Research has highlighted the importance of using Rasch models to demonstrate psychometric support beyond what is gleaned from traditional analyses (e.g., factor analysis; Smith et al., 2002). Further, the Rasch model was selected for the current study as it allows for sample-free measurement. Namely, the calibration of item difficulty does not depend on the sample of people measured, nor does the calibration of person ability depend on the sample of items administered. This is desirable as it suggests the patterns of difficulty on the GNG Screen can be generalized beyond the current sample of people and items.

The Rasch model proports that an instrument measures a single latent dimension, and that individuals and items can be organized by difficulty and ability along this underlying dimension. The model selected for completing item level scaling and validation within the current study is the dichotomous Rasch model as there is no partial scoring on items within a block; each item receives a dichotomous score of 0 (incorrect) or 1 (correct). Modeling included only "no-go" items, given the current guiding conceptual model, with response inhibition being the hypothesized dimension to underlie performance across “no-go” items. As noted above, the Rasch model includes one parameter, meaning items can vary only in terms of difficulty. The model presumes that guessing is either part of the ability or contributes random noise to the data. Further, the model ascertains that all items weigh equally on the factor, or that all items are equivalent in terms of discrimination of person ability.

A key proposition of the Rasch model is that easier items are more likely to be answered correctly by all individuals (as compared to more difficult items; Bond & Fox, 2001). People with higher levels of ability are more likely to answer all items correctly than those with lesser ability. Thus, the Rasch model proports that both items and people can be arranged from least difficult/ least ability to most difficult/ most ability. Analyses utilize this order to calibrate item difficulty values on an equal-interval sale, which is referred to as a logit scale. The logit scale is derived from a logarithmic transformation of ordinal data to interval data. On the scale, each person and each item receive a measure, in which the distance from the mean item difficulty is described in logits—referred to as “item difficulty” for items and “person ability” for people. Item difficulty values demonstrate each items placement according to the probability of the item being answered correctly by all people, along the dimension the instrument proposes to measure—in this case response inhibition.

Since the scale is calibrated using interval level data, rather than ordinal data, the distances and orders of items and people is meaningful. For example, an item with a difficulty measure of “2” is twice as hard as an item with a difficulty level of “1;” this also applies to person ability. Both difficulty and ability values can be positive or negative, with a typical range of -3 to 3 logits. The probability of a person answering an item correctly at their same level of difficulty is set to 50%. For example, the probability of a person at 2 logits of ability answering an item at 2 logits of difficulty is 50%. This probability changes based on item difficulty; the probability of the same person answering an item correctly at 1 logit of difficulty is 75% and at 3 logits of difficulty is 25%. Thus, Rasch modeling provides information on the items that are more difficult, and also quantifies the difference in difficulty. Again, the same applies for people.

Rasch Model Assumptions. Model assumptions, which include unidimensionality, conditional independence, sufficiency, and monotonicity, are detailed below and were evaluated as part of the study's primary analyses. Of note, unlike traditional hypothesis testing, the assumptions for employing the Rasch model are not characteristics of the data presumed to be true which need be assessed *a priori* (Bond & Fox, 2001). Rather, they are best conceptualized as ideals to be approximated and are inherent to the analysis itself. Failing to meet model assumptions, in part, suggest problems with the measurement of a given instrument and can be used to inform revisions.

<u>Unidimensionality.</u> Per Bond and Fox (2001), values for person ability and item difficulty are considered meaningful if the instrument measures one dimension. In other words, all non-random variance found in the data can be accounted for by a single dimension. One method for measuring dimensionality is a principal component analysis (PCA) of the Rasch residuals (Research Question 1). If a measure is unidimensional, it is expected residual factor loadings will be small and/or meaningless. Dimensionality can also be assessed by examining item fit statistics, which provide summaries of responses that differ based on what is predicted by the Rasch model (Research Question 2).

Conditional Independence. The assumption of conditional independence (also referred to as local independence) suggests that responses to one item does not depend on the success or failure of responses to separate items. In other words, the items within the instrument are related to each other only by the latent trait measured by the instrument. Conditional independence is related to unidimensionality, as dependent items may appear as separate dimensions in dimensional or factor analyses. Conditional independence can be assessed by examining the residual correlations derived from factor analysis (Research Question 1).

Sufficiency. Sufficiency is necessary for sample free measurement, meaning all information necessary for estimating person ability is included in the solved items, and all information needed for calibrating item difficulty is included in the number of times an item was solved. That is to say, person and item metrics do not depend on which people gave a correct response or which items were solved correctly by a given person. Per Linacre (1992), item fit to the Rasch model is a test of sufficiency, which is assessed in Research Question 2.

Monotonicity. Monotonicity suggests that people with higher ability correspond with a higher response probability. In other words, people with more ability on a measured trait (response inhibition) should have a higher probability of responding to an item correctly. Monotonicity is assessed by visual inspection of the expected score item characteristic curve (ICC) for each item (Research Question 2).

Research Question 1: What is the dimensionality of the GNG Screen?

Evaluating Unidimensionality. To assess the dimensionality of the GNG Screen, local independence was first assessed. This ensures all items are independent of each other and only correlate based on the latent trait they are proposed to measure (i.e., response inhibition). In addition to providing information about the way in which items relate, as noted above, local independence is a core assumption to the Rasch model. Residual correlations were derived via confirmatory factor analysis (CFA). Items with values greater than +/- 0.2 were deemed as violating local independence.

To further assess whether the GNG Screen measures a single underlying construct, dimensional analyses were conducted using Rasch modeling. Additionally, these analyses provide further evidence for the unidimensionality assumption. PCA of the residuals was conducted using Winsteps. Analyses were conducted using all items of the GNG Screen, as well as individually for each block. First, the percentage of variance explained by the dimension, as compared to the percentage of residual variance, was examined to assess dimensionality. Linacre (2005) recommends that percentage of variance explained by the dimension be at least 40%. Additionally, contrasts were examined; Linacre (2005) recommends the variance explained by the first contrast be below 5%. Lastly, eigenvalues of contrasts are examined; If the first contrast eigenvalue is small, it can be regarded as noise, while eigenvalues greater than 3 suggests systematic variance indicative of a second dimension.

Evaluating Block Factor Structure. To better understand the factor structure of the GNG Screen, EFA was used as to examine the latent trait underlying blocks. EFA is a classification of multivariate statistical methods that aims to identify the smallest number of hypothetical constructs that can parsimoniously account for variation among measured variables (Tucker & MacCallum, 1997). In other words, the goal of EFA is to identify the common factor(s) that illustrate the structure of measured variables. Factors are presumed to be unobservable characteristics, which are demonstrated in variation in performance (i.e., scores) within a given instrument. Thus, in the current study, response inhibition is the theoretical latent construct.

Prior to analysis, the factorability of the data were assessed. First, the correlations between variables were examined. Reasonable factorability was determined if each item was correlated at the 0.01 significance level with at least one other item. Bartlett's test of sphericity was used to the test the overall significance of each correlation within the matrix to further determine if it was appropriate to use the factor analytic model. Further, the Kaiser-Meyer-Olkin measure of sampling adequacy was used to indicate if the strength of the relationship was at or above the recommended value (KMO = .60; Dziuban & Shirkey, 1974). EFA using a maximum likelihood extraction method was then conducted, utilizing *Block Commission Score*, to determine the factor structure of the overall GNG Screen. Factors with eigenvalues greater than 1.0 were retained (Kaiser, 1960). Factor loadings were examined, and block performance was expected to load significantly on one factor. To further examine the factor structure, fit was examined using two- and three-fixed factor models.

Research Question 2: Do items progress in difficulty according to theoretical expectations?

Item difficulty values, fit indices, and reliability were calibrated with the Rasch model using joint maximum likelihood estimation as implemented by Linacre (2020).

Evaluating Item Difficulty. Item difficulty estimates represent the continuum of the construct as measured in logits (Bond & Fox, 2001). Based on the guiding conceptual model, it was predicted "no-go" items within a given block will have similar item difficulties (e.g., all “no-go” items within Block 1 will have similar difficulty). Additionally, as hypothesized by the guiding conceptual model, it was predicated “no-go” items with longer inter-stimulus intervals, higher discrimination, and/or lower number of distractors will be easier than items with shorter inter-stimulus intervals, lower discrimination, and/or higher number of distractors. To ensure monotonicity, the ICC for each item was examined via visual inspection. The Wright Map was also examined to determine the distribution of items.

Item difficulty values were calibrated; average item difficulty for each block was then computed. A one-way ANOVA was conducted to examine significant differences in average item difficulty between blocks. Finally, exploratory correlation analyses were conducted to examine item characteristics not described by the guiding conceptual model—including ordinal placement within a given block and number of preceding “go” items— that may relate to item difficulty.

Evaluating Item Fit Statistics. Fit indices aid in assessing whether item difficulty and person ability values can be regarded as meaningful summaries of the data. Further, these values identify both individuals and items that are not corresponding with the expected pattern of performance. The expected pattern is defined by the interaction between an individual’s ability and the item’s difficulty. Individuals with ability levels higher than that required for a particular item have a higher probability of providing a correct response. Correspondingly, individuals with lower ability than an item have a lower probability of correct response. When individuals with relatively higher ability levels fail on easier items, or when individuals with relatively lower ability levels pass on more difficult items, there is a deviation from the expected pattern of performance. This deviation is captured by fit indices. In the current study, when misfit was identified, efforts were made to determine the problem contributing to misfit, with the ultimate goal of addressing these concerns in future iterations of the GNG Screen (Wright & Stone, 1979). Linacre (2005) recommends an acceptable range of 0.5-1.5 for fit values, which was employed in the current analyses.

Evaluating Person/Item Reliability. Reliability estimates were also calculated using the Rasch model for items and individuals. The item reliability index provides a measure of how replicable the placements of item difficulty values along the logit scale would be if the same items were administered to another sample with similar levels of person ability (Bond & Fox, 2001). The person reliability index tells how replicable the placement of individuals along the logit scale would be if the same persons were administered another group of items measuring the same construct (Bond & Fox, 2001). Item and person reliability values range from 0 to 1, with values of .80 and above considered acceptable (Fox & Jones, 1998). This is different than Cronbach's alpha, which represents the repeatability of raw scores. Because of this difference, it has been argued that Rasch reliability is more conservative, and therefore possibly less misleading, than Cronbach's alpha for reliability estimates of generalizable measures.

Evaluating Floor/Ceiling Effects. Floor and ceiling effects were also examined using Rasch calibrations. Floor or ceiling effects were considered to be present if more than 15 % of respondents achieved the lowest or highest possible score, respectively. The presence of floor and ceiling effects are indicative that extreme items are missing in the lower or upper end of the scale, suggesting limited content validity.

Sample Size Requirements

The recommended sample size for analyses employing Rasch modeling is roughly 200 participants (Wright & Tennant, 1996), although item calibrations stable within +/- 1 logits can be achieved with as few as 30 participants (Linacre, 1994). For EFA, Kline (1979) recommended a minimum sample size of 100 as a general guideline. Thus, the current sample includes an adequate number of participants for analyses involving Rasch Modeling and EFA.

CHAPTER 3: RESULTS

Sample Descriptives

The total sample included 119 individuals: 13 participants in the Early Childhood group (M = 5.60 years; SD = 1.08), 10 participants in Middle Childhood (M = 7.90 years; SD = 0.61), 16 participants in Late Childhood (M = 10.60 years; SD = 0.91), 9 participants in Early Adolescence (M = 13.70 years; SD = 1.42), 19 participants in Late Adolescence (M = 18.20 years; SD = 1.36), and 52 participants in Early Adulthood (M = 25.70 years; SD = 4.00). Of note, there are 23 cases within the target age-rnage for the test, which is a notable limitation and further discussed within Chapter 4. Demographic characteristics for each group are provided in Table A.4. The total sample consisted of 58% females and 42% males; and was 80% White, 7% mixed race, 5% American Indian and Alaskan Native, 4 % Black, 3% Asian, and 1% Hawaiian or Other Pacific Islander. Data from all age groups were analyzed together to provide as diverse a range as possible in order to improve estimation capability for item calibrations within the Rasch model, as well as meet sample size requirements for EFA. Results for analyses are organized below in terms of research question.

Demographic analyses. Average *Total Commission Score* across gender and ethnicity, by age group, is reported in Table A.5. Further, Average *Block Commission Score* across each block is reported by age group in Table A.6. A two-way ANOVA was conducted to examine the effects of gender and age group on *Total Commission Score*, which is reported in Table A.7. The interaction between gender and age group on *Total Commission Score* was not significant ($F(5, 107) = 0.519$, $p = 0.76$), nor was the main effect of gender ($F(1, 107) = 0.460$, $p = 0.50$). However, results suggested a significant main effect for age group ($F(5,107) = 9.17$, $p < .001$). All pairwise comparisons were run, where reported p-values are Bonferroni-adjusted; pairwise comparisons are reported in Table A.8.. Early childhood was associated with a *Total Commission Score* 17.88 points lower than Late Childhood, 22.68 points lower than Early Adolescence, 24.31 points lower than Late Adolescence, and 26.28 points lower than Early Adulthood. No other significant differences between groups were observed.

Averages for *Block Omission Score, Block PCT*, *Block RT*, *Block RTV,* and *Block Efficiency* are reported by age group in Tables A.9, A.10, A.11, A.12, and A.13, respectively. Additional exploratory two-way ANOVAs were conducted to examine the effects of age and gender on *Total Omission Score, Total PCT, Total RT, Total RTV,* and *Total Efficiency.* Results are reported in Table A.14. The interaction between age group and gender across metrics was insignificant, as was the main effect of gender. Though, significant main effects for age group were found across all metrics (*Total Omission Score*: $F(5,107) = 3.25$, $p = .009$; *Total PCT:* $F(5,107) = 5.14$, $p < .001$; *Total RT*: $F(5,107) = 7.83$, $p < .001$; *Total RTV*: $F(5,107) = 4.74$, $p = 0.001$; *Total Efficiency*: $F(5,107) = 10.47$, $p < 0.001$).

All pairwise comparisons were run, where reported p-values are Bonferroni-adjusted; pairwise comparisons are reported in Tables A.15, A.16, A.17, A.18, and A.19. Early childhood was associated with a *Total Omission Score* 29.02 points lower than Late

Adolescence and 31.74 points lower than Early Adulthood. Early Childhood was 22.96% less accurate than Early Adolescence, 22.46% less accurate than Late Adolescence, and 23.96% less accurate than Young Adults. Those in Early Childhood were .36 seconds slower than those in Early Adolescence .29 seconds slower than those in Late Adolescence, and .32 seconds slower than those in Young Adulthood. Those in Middle Childhood were .30 seconds slower than those in Early Adolescence and .26 seconds slower than those in Early Adulthood, while those in Late Childhood were .21 seconds slower than those in Early Adulthood. Those in Early Childhood demonstrated significantly more variable response times than those in Young Adulthood. Lastly, those in Early Childhood were less efficient in responding than those in Early Adolescence, Late Adolescence, and Early Adulthood. Those in Middle Childhood were significantly less efficient in responding than those in Early Adolescence, Late Adolescence, and Early Adulthood. Those in Late Childhood were significantly less efficient in responding than those in Early Adulthood. No other significant differences between age groups were observed across exploratory metrics.

Correlations Across Test Metrics. To examine the relationship between performance across metrics, Pearson's correlation coefficients were computed between *Total Commission Score*, *Total Omission Score*, *Total RT*, and *Total RTV*. The remaining metrics (*Total PCT* and *Total Efficiency*) were not included in correlation analyses due to their composite nature. After correction for multiple analyses, there were statistically significant, negative correlations between *Total Omission Score* and *Total RT* ($r = -0.78$, $p < .001$), as well as *Total RTV* ($r = -0.78$, $p < .001$), illustrating as accuracy on "go" items increased, time to respond and variability in response time both decreased.

Further, there was a statistically significant, positive correlation between *Total RT* and *Total RTV* (r = 0.88, p < .001), suggesting that as time to respond increased, so did variability in response time. Of note, *Total Commission Score* was not significantly correlated with other metrics. Refer to Table A.20 for a complete correlation matrix.

Research Question 1: What is the factor structure of the GNG Screen?

Evaluating Unidimensionality. The local independence of "no-go" items within the GNG Screen was assessed to establish their independence from each other, such that items only correlated in terms of the latent trait that they were proposed to measure (i.e., response inhibition). No residual correlations were above +/-0.2, indicating the assumption of local independence was met.

Rasch modeling was used to assess the dimensionality of the GNG Screen (via PCA); results are reported in Table A.21. When examining the entire GNG Screen, results showed 37.6% of variance explained by the measure, with the first contrast accounting for 3% of the unexplained variance. Thus, when investigating the entire test, dimensionality results approached meeting the first criterion (i.e., at least 40% explained by the measure) and met the second criterion (i.e., less than 5% explained by the first contrast) set by Linacre (2005). Of note, there was more unexplained variance than explained variance; however, the eigenvalue of the first contrast was less than what would indicate a second dimension (i.e., < 3).

When examining individual blocks within the GNG Screen, no blocks met the first criterion for unidimensionality set forth by Linacre (2005; at least 40% variance explained by the measure). Variance explained by the measure ranged from 20.9% (Block 3) to 38.3% (Block 6). Likewise, no blocks met the second criterion provided by

Linacre (2005; less than 5% of variance explained by the first contrast of the residuals). Percent of variance accounted for by the residuals ranged from 11.9% (Block 5) to 15.7% (Block 1). Regardless, similar to findings from the overall test, eigenvalues of the first contrast across blocks were less than what would indicate a second dimension (i.e., < 3). Refer to Table A.22 for a comprehensive overview of results across blocks.

Evaluating Block Factor Structure. To further examine the factor structure of the GNG Screen, an EFA was run using *Block Commission Score* across all blocks. The suitability of EFA was assessed prior to analysis. Inspection of the correlation matrix indicated that all variables had at least one correlation coefficient at the .01 significant level. The overall KMO measure was .89 with individual KMO measures all greater than .80. Further, Bartlett's Test of Sphericity was statistically significant, indicating the data were likely factorizable. Factor loadings are reported in Table A.23 and model fit statistics are reported in A.24. EFA revealed one factor had an eigenvalue greater than one, which explained 71.10% of the total variance. Visual inspection of the scree plot indicated this factor should be retained.

To determine if the one factor model did possess optimal fit, additional EFAs were run using two- and three-fixed factor models. Analyses revealed poor model fit as compared to the one-factor model, with eigenvalues for the second and third factor falling below what would indicate the presence of two or three dimensions (*two-factor model*: factor 1 eigenvalue = 5.05, factor 2 eigenvalue = 0.18; *three-factor model*: factor 1 eigenvalue = 5.05, factor 2 eigenvalue = 0.18; factor 3 eigenvalue = 0.15). Factor loadings are reported in Tables A.25 and A.26.

Research Question 2: Do items progress in difficulty according to theoretical expectations?

Evaluating Item Difficulty. Visual evaluation of ICC curves across all items indicated monotonicity, as the expected ICC approximated the Rasch model-predicted ICC. The Wright Map was also examined using visual inspection to determine the distribution of items. As illustrated in Figure A.1, pronounced groupings (i.e., overlap) of items are present, with a large portion of ability falling above the plotted range of difficulty.

Table A.27 provides average item difficulty, reported in logits, for "no-go" items within each block. Visual inspection indicated that blocks progressed in the following order of difficulty, from least to most difficult: Block 1 (*M* = -3.72 , *SD* = 0.65), Block 2 (*M* = -3.63 , *SD* = .74), Block 3 (*M* =3.10 , *SD* = 0.71), Block 4 (*M* = -2.82 , *SD* = 0.93), Block 5 (*M* = -2.75, *SD* = 0.83), Block 7 (*M* = -2.36, *SD* = .56), Block 6 (*M* =-2.22 , *SD* = 0.66), Block 8 (*M* = -2.14, *SD* = 0.30). A one-way ANOVA demonstrated significant differences between blocks in terms of average item difficulty ($F(7, 64) = 5.87$, $p = <0.001$; see Table A.28). Tukey post-hoc analysis, which is reported in Table A.29, revealed Block 1 was significantly easier than Blocks 6, 7, and 8. Additionally, Block 2 was significantly easier than Blocks 6 and 8. No other blocks differed significantly in terms of average difficulty. Although average difficulty per block generally progressed in order of theoretical expectations, visual inspection of difficulty values indicated variability within each block, with overlap in item difficulty between blocks.

To better understand additional potential factors related to item difficulty, characteristics of items beyond that explained by the guiding conceptual model—

including ordered placement within the block and number of preceding “go” items—were examined. Pearson’s correlation coefficients indicated number of “go” items preceding a given “no-go” item was positively correlated with item difficulty, in logits (p = .002).

Evaluating Item Fit Statistics. Fit indices were examined to determine unidimensionality of “no-go” items within each block. Table A.30 provides average infit and outfit mean-squares for all blocks. Average infit for each block fell within the 0.5 to 1.5 range deemed acceptable by Linacre (2005), thus indicating acceptable fit to the Rasch model. In addition to average mean-squares for each block indicating acceptable fit—which represent the mean of mean-square values for all commission items in a given block—all individual items within each block also showed infit values within the acceptable range. In terms of outfit, the average mean-square for Block 3 fell outside the acceptable range (M = 1.6), while the average outfit values for the remaining blocks fell within the 0.5 to 1.5 range. According to individual outfit values, all blocks contained misfitting items, with Blocks 7 and 8 containing the fewest (1 item, for each) and Block 3 containing the most (5 items).

Evaluating Person/Item Reliability. The person reliability index was calculated, with a value at .80 or above considered acceptable. The GNG Screen showed an acceptable person reliability index of .86. Item reliability indices were also calculated, again with values at .80 or above considered acceptable. The GNG Screen showed an acceptable item reliability of .83.

Evaluating Floor/Ceiling Effects. As there are 9 “no-go” items per block, with a minimum score of 0 and a maximum score of 1, the lowest possible sum of scores per block is 0 and the highest possible sum of scores per block is 9. Floor or ceiling effects were considered present if more than 15% of participants achieved the lowest or highest possible score on a given block, respectively (Linacre, 1997). When examining the entire sample, ceiling effects were observed across all blocks. That is, 64.7% of participants demonstrated the highest possible score on Block 1, 58.0% on Block 2, 49.6% on Block 3, 43.7% on Block 4, 46.2% on Block 5, 47.1% on both Blocks 6 and 7, and 52% on Block 8. No blocks exhibited floor effects when examining the entire sample. Given the target age range of the GNG Screen, potential floor and ceiling effects were also examined including only participants in Early Childhood (ages 4 - 6 years) and Middle Childhood groups (ages 7 – 8 years). When examining this subsample, mild ceiling effects were observed across Blocks 1 and 2, on which 17.4% of participants demonstrated the highest possible score for both blocks. Again, no blocks exhibited floor effect.

CHAPTER 4: DISCUSSION

Previous research has demonstrated a strong relationship between symptoms of ADHD and academic underachievement (e.g., Fraizer et al., 2007; Hinshaw, 1992; Rapport, Scanlan, & Denney, 1999) and interventions specific to academic deficits in children with ADHD are available (DuPaul et al., 2004). These interventions are most effective if implemented early, before secondary concerns compound the primary observed deficits (DuPaul & Kern, 2011; Sonuga-Barke, 2010). Performance based screening is one method for determining need for early intervention; however, extant measures of attention have psychometric, theoretical, and practical limitations for the purpose of large-scale screening. The current study proposed that measures of response inhibition have utility in the screening of attention deficits. Specifically, this study evaluated a novel instrument that measures response inhibition via a go/no-go paradigm—the GNG Screen— which addresses limitations of extant measures. This instrument was organized by a theoretically-driven, guiding conceptual model and administered through a fully automatized format. The current study evaluated the guiding conceptual model that organizes the measure in terms of item difficulty, as well as examined the reliability and dimensionality of the instrument. Results for each research question will be discussed below, followed by a summary of conclusions, limitations and directions for future research, and implications for practitioners.

Discussion of Results

Demographic analyses. The current study examined the effects of demographic characteristics, including age and gender, on test performance. The interaction between gender and age group (i.e., Early Childhood, Middle Childhood, Late Childhood, Early Adolescence, Late Adolescence, and Early Adulthood) on *Total Commission Score* was assessed via two-way ANOVA. No significant interaction was found, suggesting the effect of age group on *Total Commission Score* was the same for both males and females. Moreover, no significant difference was found between males and females. Though, consistent with extant literature, a significant main effect was found for age group. Namely, significant differences in *Total Commission Score* were found for participants in the Early Childhood group and participants in the Late Childhood, Early Adolescence, Late Adolescence, and Early Adulthood groups. No other significant differences were found across age groups. Due to restrictions in sample size, the effect of race on *Total Commission Score* was not assessed in the current study, which is a notable limitation and is further discussed below (see Study Limitations and Future Directions).

Findings from demographic analyses align with the developmental nature of EF. More specifically, the pattern of results reflect observed "critical periods" for the development of inhibitory control (Barkley, 2012). Primitive elements of EF emerge throughout early childhood (i.e., the first 5 years of life, approximately), creating a fundamental foundation for the growth of higher-order cognitive processes during adolescence and adulthood. Inhibitory control emerges around the age of 4 years, and rapidly develops between the ages of 5 and 11 years. While inhibitory control and EF at large continue to develop through late childhood and into early adulthood, the rate of development is less rapid, accounting for non-significant results later in development when examining differences by age groupings.

Non-significant findings regarding the effect of gender on *Total Commission Score* is generally unsurprising given limited support for sex differences in go/no-go performance, with a wealth of literature demonstrating no significant differences (e.g., Erickson et al., 2005; Li, Zhang et al., 2009; Thakkar et al., 2014). Though, boys and girls have been found to perform differently on go/no-go tasks at various points throughout child (Hasson & Fine, 2012); however, the current study likely lacked statistical power to detect subtle interactions. Further, recent studies examining go/no-go performance in healthy adults have suggested mild sex differences, in that adult females generally outperform adult males (Sjoberg & Cole, 2018). However, given observed ceiling effects of the GNG Screen, as well as the overall target age-range of the current measure, it is unsurprising the instrument did not detect potential sex differences at the upper end of ability.

For exploratory purposes, the interaction between gender and age group across secondary metrics gleaned from the GNG Screen was also assessed via two-way ANOVA (i.e., *Total Omission Score*, *Total PCT*, *Total RT*, *Total RTV*, and *Total Efficiency*). No significant interactions were found across metrics, suggesting the effect of age group was the same for both males and females. Moreover, no significant differences between males and females were found across metrics. Similar to *Total Commission Score*, non-significant findings are unsurprising for gender differences. Though, again consistent with *Total Commission Score*, significant main effects were found for age group across all metrics, which are described below.

Omission Scores: Current analyses revealed significant differences in *Total Omission Score* for participants in the Early Childhood group and participants in the Late Childhood, Early Adolescence, Late Adolescence, and Early Adult groups. No other significant differences in *Total Omission Score* were found across age groups. Current findings align with developmental expectations for errors of omission, which are a measure of sustained attention (Barkley, 2012). Similar to inhibitory control, sustained attention is proposed to rapidly develop during early childhood, with continued development throughout adolescence and into early adulthood, albeit less rapid.

Percentage of Correct Trials: PCT is a composite score, accounting for overall accuracy across all "go" and "no-go" items. Thus, theoretically, it is a pooled measure of both sustained attention and response inhibition. Previous research has not examined potential changes in such composite scores across developmental periods; however, unsurprisingly, results were generally consistent with those of *Total Commission Score* and *Total Omission Score*, in that participants in the Early Childhood group were less accurate than those in Early Adolescence, Late Adolescence, and Early Adulthood groups.

Reaction Time, Reaction Time Variability, and Efficiency: Limited meaningful findings have been found regarding the relationship between RT and attention deficits; however, RTV and efficiency— which are measures of information processing and considered core EFs in the present study— have been repeatedly associated with attentional deficits (Barkley, 2012). In the current study, those in the Early Childhood group were significantly slower to respond across trials than those in the Early Adolescence, Late Adolescence, and Early Adulthood groups. Moreover, those in Middle

Childhood were slower to respond than those in Early Adolescence and Early Adulthood, and those in Late Childhood were slower to respond than those in Early Adulthood. Regarding RTV, those in the Early Childhood group demonstrated significantly more variable response times than those in Early Adulthood. Lastly, in terms of efficiency, those in Early Childhood were less efficient in responding than those in Early Adolescence, Late Adolescence, and Early Adulthood; those in Middle Childhood were significantly less efficient in responding than those in Early Adolescence, Late Adolescence, and Early Adulthood; and those in Late Childhood were significantly less efficient in responding than those in Early Adulthood.

Reaction time has long been demonstrated to improve with age during early life (Surwillo, 1977), and similarly, information processing—including RTV and efficiency—has been demonstrated to follow the observed developmental trajectory of other core EFs (Barkley, 2012). Thus, findings across *Total RT*, *Total RTV*, and *Total Efficiency* are consistent with previous literature.

Correlations Across Test Metrics. As previously noted, commission scores are the primary foci of the current study, due to the established relationship with response inhibition (Barkley, 2012). However, other aspects of attention and EF are captured within a go/no-go paradigm, including sustained attention and information processing. Thus, the relationship between *Total Commission Score*, *Total Omission Score*, *Total RT*, and *Total RTV* was assessed. Results from the present analyses suggest that *Total Commission Score* was not significantly correlated with any secondary metric. This result is unexpected, as although these metrics are thought to measure distinct EFs, individual functions tend to vary at similar levels *within* people (Barkley, 2012). It is unclear why the current pattern is observed; however, it may be explained, at least in part, by the age-range of the current sample and previously discussed trends observed within the development of EF. The relationship between *Total Commission Score* and extant validated measures of EF should be further examined in future research, as discussed below. Though, consistent with what is known about EF, *Total Omission Score* was negatively correlated with *Total RT* and *Total RTV*, in that those with higher omission scores (i.e., less omission errors; increased sustained attention) demonstrated less time to respond (i.e., faster), as well as less variability (i.e., more consistent; Barkley, 2012). Unsurprisingly, *Total RT* was positively correlated with *Total RTV*, suggesting those with faster response times were also less variable in their responses.

Research Question 1: What is the dimensionality of the GNG Screen?

Evaluating Unidimensionality. The local independence of "no-go" items in the GNG Screen was assessed to establish that items 1) are independent from each other, and 2) only correlate in terms of the latent trait they are proposed to measure (i.e., response inhibition). No residual correlations were above +/-0.2, demonstrating the assumption of local independence was met. Local independence is related to unidimensionality in that dependent items may appear as separate dimensions in factor or dimensional analyses. Thus, the absence of dependent items suggests the way in which GNG Screen items relate is likely not contributing to observed noise in dimensional analyses, described below.

Regarding unidimensionality, analyses revealed the GNG Screen (all items together) approached Linacre's (2005) guidelines, with 37% of variance explained by the measure and 3% of variance explained by the first contrast. Importantly, the dimensionality parameters (via examination of contrast eigenvalues) indicated the GNG Screen does not have a second dimension; that is, the GNG Screen is unidimensional. In addition to exhibiting the instrument meets the unidimensionality assumption, this finding also suggests there are no additional factors being measured by the GNG Screen—which is important for demonstrating construct validity. Findings further suggest that unexplained variance is too random to form a second dimension. As such, unexplained variance may be indicative of items not contributing toward the overall construct, rather than multidimensionality. Results from dimensionality analyses further indicated individual blocks did not meet Linacre's criteria for adequate unidimensionality (2005), though parameters again suggested that all blocks did not have a second dimension.

Evaluating Block Factor Structure. Results from EFA provide further evidence for the unidimensionality of the GNG Screen. Namely, initial analyses demonstrated all blocks loaded onto one-factor, providing excellent fit to the data. This evidence suggests the current GNG Screen includes one latent factor, which aligns with the theoretical model. Additional EFAs using two- and three- fixed factor models indicated poor fit, providing additional support for the unidimensionality of the measure.

Interpretation of Dimensionality Findings. When interpreting dimensionality results, it is important to consider the multidimensional nature inherent to the construct of EF (Barkley, 2012). Many studies have demonstrated EFs are difficult to measure in isolation (see Anderson, 2002) . Indeed, efforts were made to reduce the impact of distinct functions, such as cognitive flexibility and working memory, in order to create a "pure" measure of response inhibition. For instance, blocks were reordered from the initial theoretical hierarchy of difficulty to eliminate switching between rules for responding, and ultimately reduce the load on cognitive flexibility. Despite efforts to target inhibitory control within the GNG Screen, aspects of separate EFs likely remain within the measure. For example, increased working memory load was presumably present in blocks with increased distractors, as accurate responding required holding in memory multiple targets/non-targets. Thus, commission errors in Blocks 3, 4, 7, and 8 could be driven, at least in part, by lapses in working memory, as opposed to errors in inhibitory control. Further, per Barkley (1997), inhibition encompasses three interrelated components: response inhibition, action cancellation, and interference control. Thus, in addition to the difficulty of measuring EFs in isolation, it is probable the interrelated components of inhibition are captured by the instrument, adding noise to the data.

Beyond the inherent difficulty in the measurement of EF, it is imperative to consider the fundamental characteristics of go/no-go tasks when interpreting dimensionality findings. The essence of the task is the building of a prepotent response based on the ratio of "go" to "no-go" items. That is, presenting a single "no-go" item—without the presence of "go" items— would not provide meaningful information on a person's ability. As discussed in detail when reviewing results from Research Question 2, exploratory analyses demonstrated the number of "go" items preceding a "no-go" item significantly correlated with item difficulty. Although "no-go" items demonstrate local independence— such that the likelihood of responding to one item does not depend on the success or failure of responses to separate items—composite scores across multiple items likely provide increased information on a person's level of inhibitory control. In addition to the interpretation of dimensionality results, this finding is important to consider when determining which metrics to utilize when developing clinical cut-offs for screening purposes.

The developmental nature of EF—and more specifically response inhibition—further compounds dimensionality findings. As a person develops, EFs become more plentiful and complex, building upon foundational EFs. While a review of the relationship between higher-order EFs is beyond the scope of the current study given the target age-range of the GNG Screen, it is noted that noise within measurement of EFs increase as a function of increased complexity (Barkley, 2012). Thus, the GNG Screen may approximate a "pure" measure of response inhibition for the intended age group; however, as inhibitory control stabilizes in later development and more complex EFs emerge, variability in performance may be better explained by differing levels of separate EFs. Limitations in sample size prevent exploration of dimensionality within the intended age group alone; however, as discussed below, this age group should be a primary target of future research.

Research Question 2. Do items progress in difficulty according to theoretical expectations?

Examining Item Difficulty. As illustrated by the Wright Map, pronounced groupings of items are present, with a large portion of ability falling above the plotted range of difficulty. When exploring average item difficulty by block, blocks were found to progress in the following order, from least to most difficult: Block 1, Block 2, Block 3, Block 4, Block 5, Block 7, Block 6, Block 8. Multiple aspects of this hierarchy align with the guiding conceptual model, with minor deviations. Of note, while not all differences were statistically significant (see below), difficulty ranking is discussed to inform future iterations of the GNG Screen.

In terms of results consistent with theoretical expectations, blocks in Level 2 (i.e., Low Discrimination; Blocks 5, 6, 7, and 8) were consistently more difficult than blocks in Level 1 (i.e., High Discrimination; Blocks 1, 2, 3, and 4). When examining interstimulus interval throughout Level 2, blocks with short interstimulus intervals (i.e., faster presentation; Blocks 6 and 8) were consistently more difficult than those with long interstimulus intervals (i.e., slower presentation; Blocks 5 and 7). Further, short interstimulus interval was more difficult, irrespective of number of distractors, as theorized (i.e., Block 6 more difficult than Block 7). Additionally, blocks with higher number of distractors (Blocks 7 and 8, respectively) were more difficult than corresponding blocks with lower number of distractors (Blocks 5 and 6, respectively), as expected. Regarding number of distractors within Level 1, blocks with more distractors (Blocks 3 and 4) were found to be more difficult than those with less distractors (Blocks 1 and 2), again consistent with the theoretical model.

By comparison, when examining interstimulus interval in Level 1, short interstimulus interval was *not* consistently more difficult than long interstimulus interval. Namely, Block 3 (long interstimulus interval, high number of distractors) was found to be more difficult than Block 2 (short interstimulus interval, low number of distractors). Of note, multiple out-fitting items were found within Block 3. Thus, while deviations from the guiding conceptual model may suggest block characteristics affect difficulty in a way that is inconsistent with the guiding model and/or are representative of a distinct function (e.g., working memory), it is also likely misfitting characteristics of Block 3 are contributing to observed inconsistencies. Regardless, Block 3 should be further examined and revised before including in future iterations of the GNG Screen.

Differences in block difficulty were further investigated via one-way ANOVA; meaningful differences in average item difficulty for Block 1 and Blocks 6, 7, and 8, as well as Block 2 and Blocks 6 and 8. No other blocks differed significantly in terms of item difficulty. When examining item characteristics between blocks with significant differences, findings suggest interstimulus interval and level of discrimination may provide the most change. These item characteristics should be considered if an increased range of difficulty is desired within future iterations of the GNG screen. Moreover, results suggest a cumulative effect of item characteristics, such that changing one aspect leads to insignificant change, while altering multiple item characteristics leads to meaningful change. Altogether, findings support a reduction in the number of blocks—and ultimate reduction in test length—within future versions of the GNG Screen. This outcome is particularly salient, as brevity is a critical aspect of screening measures (DuPaul & Kern, 2011).

While results at large align with the guiding conceptual model, item difficulty varied within each block. Exploratory analyses revealed an additional item characteristic—number of "go" items preceding a given "no-go" item—as related to difficulty. That is, difficulty increased as a function of increased preceding "go" items. This is unsurprising given the core assumption of go/no-go tasks: the measurement of inhibitory control through the building of a prepotent response based on the "go" to "no-go" ratio (Barkley, 1997; Berlin & Bohen, 2002). In line with previous measures utilizing a go/no-go paradigm, "no-go" items were ordered at random throughout each block within the GNG Screen (e.g., Berlin & Bohen, 2002). Of note, however, preceding "go" items varied within and across blocks. When refining later versions of the GNG Screen, efforts should be made to ensure an even distribution of "go" items proceeding "no-go" items across blocks, while also maintaining the otherwise random order of presentation.

In summary, results are generally consistent with the guiding conceptual model prediction that shorter inter-stimulus interval, decreased item discrimination, and increased number of distracters correspond to increases in item difficulty, in addition to number of preceding "go" items. These factors are parameters that can be manipulated to alter difficulty and ensure included blocks range in difficulty. That is, results provide empirical evidence for performance in that an individual should perform better on blocks with longer inter-stimulus interval, increased item discrimination, and decreased number of distractors. Moreover, an individual should begin to struggle as items progress towards shorter inter-stimulus intervals, decreased item discrimination, increased number of distractors, and increased preceding "go" items, and therefore increase in difficulty.

Examining Item Fit. In addition to item difficulty, fit indices—which are indicators of unidimensionality at the item level —were examined for items within each block. Across all blocks, average infit mean-squares fell within the acceptable range. Conversely, average outfit mean-squares fell within the acceptable range for 7 of the 8 blocks (i.e., Block 3 fell outside of the acceptable range). According to individual item infit values, no blocks contained misfitting items. However, when examining item level outfit values, all blocks contained individual misfitting items (20 out of 72 total "no-go" items).

Of note, item outfit is an unweighted metric, indicating it is influenced by outlying scores. By comparison, infit is weighted more strongly for people whose scores fall near the item difficulty value for a particular item. According to Bond and Fox (2001), since infit is a weighted metric, it likely provides more meaningful information regarding the performance of an item (as compared to outfit). Thus, when considering item unidimensionality according to infit indicators, results demonstrate all items are unidimensional (i.e., each item measures one underlying construct). This provides additional psychometric support for the unidimensionality of the GNG Screen.

Examining Reliability. The overall item reliability of the GNG Screen was acceptable (>.80). Further, the person reliability, an IRT equivalent for internal consistency, was also acceptable (>.80). These findings provide important psychometric support for the GNG Screen as they demonstrate the replicability of items and people, respectively. Namely, the item reliability index demonstrates the placements of item difficulty values along the logit scale would be largely similar if items were administered to another sample with similar levels of person ability. Further, the current person reliability index tells us the placement of individuals along the logit scale would be largely similar if people were administered another group of items measuring the same construct.

Examining Floor and Ceiling Effects. Despite all items showing acceptable infit—as well as the instrument, as a whole, demonstrating acceptable item reliability—analyses revealed ceiling effects across multiple blocks. Specifically, when examining the entire sample, ceiling effects were observed across all blocks, with 43.7 to 58.0% of individuals obtaining the maximum score. When including only participants in the Early and Middle Childhood groups, mild ceiling effects remained for Blocks 1 and 2. This is notable considering the target age range of the GNG Screen. The presence of ceiling effects suggests Blocks 1 and 2 may not be providing additional information about a person's inhibitory control, as these blocks show poor ability to discriminate among people at the upper end of ability. This discovery has direct implications for future test development; floor and ceiling effects should be monitored when replicating current findings using a larger sample of children to determine need for further revising and/or omitting additional blocks. Moreover, the presence of ceiling effects has implications for interpreting results of the present analyses, as blocks showing ceiling effects may have skewed the current results.

Summary of Conclusions

Results from the present study offer foundational support for a novel measure of inhibitory control—the GNG Screen— and inform ongoing development of the instrument. Specifically, performance on the GNG Screen followed established age-related trends in the development of EF. Analyses demonstrated emerging unidimensionality of the instrument, with important considerations for improving dimensionality in future iterations. Notably, dimensionality findings from the current study align with previous evidence suggesting EFs are difficult to measure in isolation. Regardless, replicating analyses with a more targeted age range, as well as eliminating redundant and/or outfitting blocks should improve dimensionality findings. Lastly, item difficulty values gleaned from Rasch analyses generally support the guiding conceptual model; however, examination of differences in difficulty suggests a reduction in length (i.e., fewer blocks) may be sufficient for capturing the same range of difficulty.

More specifically, the following revisions at the item and block level should be made in the next version of the GNG screen: According to difficulty analyses, blocks from both Level 1 and Level 2 should be retained, particularly Blocks 1, 2, 6, 7, and 8. Of note, this will reduce the test length from approximately 18 minutes to 8 minutes. While redundancy in item difficulty will likely remain, additional revisions should be made after replicating the current analyses with sample comprised exclusively of the target age group. Importantly, the aforementioned combination of blocks does not include Block 3, which demonstrated significant item outfit per Rasch analyses. In terms of ceiling and floor effects, Blocks 1 and 2 demonstrated mild ceiling effects when examining the subsample of participants in Early and Middle Childhood groups. However, to ensure an adequate range of difficulty at the lower end of ability, these blocks should be retained within the next iteration. If ongoing ceiling effects and/or item redundancy is observed, the GNG Screen should be further scaled down. These changes, in addition to a more focused sample of participants, will likely increase the unidimensionality of the measure. If ongoing noise in dimensionality analyses is observed, the order of item presentation within each block should be revised to account for number of "go" items preceding "no-go" items. Additional changes driven by software limitations are discussed in the following sections.

Limitations and Directions for Future Research

Though the present study provides foundational empirical and psychometric support for the validity of the GNG Screen, multiple limitations may have impacted results and should be addressed in future research. Specifically, limitations were present within the sample employed, as well as characteristics of the measure and method of administration. Each of these domains, as well as associated implications for future research, are discussed below.

Sample Limitations. Several limitations of the employed sample may have impacted results and should be addressed in future research. Namely, although the sample size of 119 participants provided adequate statistical power for EFA (Kline, 1979), and was minimally sufficient for Rasch analyses (Linacre, 1994), a larger sample would increase stability of Rasch calibration for item difficulty and person ability values, as well as improve estimation with EFA models. Additional psychometric properties, such as test-retest reliability, should also be examined with a larger sample.

Of particular importance, the current sample was restricted by age distribution. While ages ranged from 4 to 31 years, the majority of participants fell in the Early Adulthood age group (ages 20 +; $n = 52$). Given the targeted age-range of the GNG Screen (i.e., ages 4 to 8 years)— in conjunction with the developmental nature of executive functioning (Barkley, 2012)— sizable ceiling effects were present when examining participants from the entire sample. Additional research with a larger, more focused sample will better inform the adequate number of blocks, characteristics of blocks, and test length for the target sample. Further, this research would improve stability of Rasch calibrations by providing increased variability in performance at the lower limits of inhibitory control. Since the GNG Screen is ultimately intended for use with young children, additional testing with this age group is necessary to further evaluate practical aspects of feasibility (e.g., user interface), in addition to psychometric properties. Lastly, future research with larger samples could establish a normative database for young children to aid in score interpretation.

Future studies should also attempt to recruit an ethnically diverse and representative sample, since the current study included mostly White participants across all age ranges. Of note, analyses exploring effects of ethnicity on performance across test metrics were not conducted due to power limitations. It is imperative to further explore ethnicity using future samples, particularly when developing normative cut-offs for clinical use. Of note, minority populations tend to be over-diagnosed and/or misdiagnosed with a variety of mental health disorders, including ADHD (see Graham-Lo Presti, Williams, & Rosen, 2019). Careful consideration regarding clinical cut-offs for performance, followed by evidence-based practices for comprehensive assessment, can help ensure both the early and accurate detection of attentional deficits. In addition to ethnicity, future standardization and norm development for the GNG Screen should include a sample that is representative in terms of gender and geography to better approximate the demographic composition of the United States. This is particularly important, as the GNG Screen is designed to be widely accessible throughout the country.

Lastly, it is imperative research includes a focus on clinical samples, which is particularly important given the practical aim of the measure. Studies should explore the differences in performance across metrics of the GNG Screen between clinical samples and typically developing children. Such studies should ultimately determine the instrument's ability to discriminate children with ADHD from those who are typically developing, in addition to differential diagnoses (e.g., SLD). Moreover, as this task is aimed to detect early executive deficits associated with a later diagnosis of ADHD, longitudinal studies designed to assess the predicative validity of this measure should be considered. Lastly, when the GNG Screen is finalized and re-evaluated, additional data should be collected to determine appropriate screening cut-offs by age and gender. Of note, while the current sample was relatively equal across gender, gender differences in both the diagnosis and presentation of ADHD have been well established (see Rucklidge, 2010). Normative studies should include both males and females with ADHD, regardless of presentation type.

Measure limitations. Multiple aspects of the software should be revised in future iterations of the GNG Screen to increase accessibility and usability. Of note, the current version is only compatible with Windows devices, which significantly restricted recruitment for the current study and will considerably limit future dissemination for clinical use. It is necessary to revise later versions for compatibility with Linux and IOS operating systems. Additionally, a version of the test that can be administered completely online, without the need for software download, should be considered to extend access for administration on tablets, cellphones, and other devices on which the software cannot be readily downloaded.

Further, as discussed previously, the length of the GNG Screen should be reduced for screening purposes. Current analyses indicated a reduction in blocks will capture the same range of difficulty; though, this should be further evaluated with larger samples of children within the target age-range. Additional considerations regarding the instrument design should be evaluated and revised in future test versions, such as controlling for potential aberrant responding. Moreover, the current study provided instructions only via recorded audio; later versions should consider including subtitles to increase accessibility. Lastly, in the current study, raw scores were converted to block scores and total scores; composite scores, such as efficiency were calculated by hand. Future iterations of the software should automatically compute these scores to be accessed electronically immediately upon completion of the task.

Of note, blocks were presented in the same order to all participants within the current study. Since block order was not randomized, the effect of block order could not be examined. While for clinical use, it is important items be administered in ascending order of difficulty, potential changes in difficulty based on block ordering is a notable instrumentation effect which requires further examiner. While the current study suggested blocks progress in general order of proposed difficulty, as outlined within the guiding conceptual model, it is possible that difficulty may be affected by block order, at least in part. In other words, for example, it is possible later blocks demonstrated increased difficulty as a product of placement, in addition to factors outlined in the current model. Thus, it imperative to examine this and other potential instrumentation effects that may threaten the validity of the instrument prior to finalization.

The current study was unable to control for several aspects of administration and environmental confounds due to the virtual nature of data collection. First, the type of device used for administration was not controlled. Future studies should examine differences between performance on touch screen, track pad, key-press, and mouse. Other device characteristics should be considered, such as screen size. The current study also did not control for the way in which participants interacted with the test, such as which hand/finger was used to provide response, distance from the screen a participant's hand was kept between trials, etc. To better control for the potential impact on reaction time, future iterations of the measure should consider prompting individuals to use the index finger of their dominant hand, and to hold their index finger at a designated area between trials. Further, environmental confounds such as presence of distractions and assistance from other individuals, were not controlled. Later iterations of the GNG Screen should be examined within more controlled environments prior to dissemination for clinical use. Lastly, given the ultimate practical aim of the measure, future studies should explore performance within large-scale screening administration to further determine the feasibility of the GNG Screen in this regard.

Additional directions for future research. In addition to research driven by limitations of the current study, there are several further areas of investigation indicated from the present study. First, the blocks and items flagged for review due to inadequate fit should be reviewed, replaced/removed, and re-evaluated. New blocks may also be created based on difficulty parameters, particularly to include an appropriate range of items for the target age range. Once a final version of the task is created, computer adaptive testing algorithms and modifications of the GNG Screen should be examined. Future studies should explore criterion validity (i.e., correlation with other measures) and predictive validity (e.g., prediction of academic course grades). Lastly, to improve accessibility throughout the country, researchers should consider developing versions of the GNG Screen available in languages other than English.

Implications for Practice & Final Thoughts

The GNG Screen has theoretical and practical advantages over currently available attention screeners. Namely, it is organized by a guiding conceptual model, is completely automated, and is administered online. The current study offers preliminary support for the validity of this novel instrument. The automated, online format provides multiple benefits for clinicians. It can be administered to large groups of individuals simultaneously without the need for a highly trained examiner, reducing resources such as training, as well as implementation time and cost. Further, the online availability of the GNG Screen can be accessed anywhere, as long as there is an internet connection available. Once finalized, schools can use the instrument to identify children who would benefit from early intervention, ultimately improving both the accurate assessment and outcomes for children with ADHD and related concerns.

REFERENCES

Alloway, T. P. (2007). *Automated Working: Memory Assessment: Manual*. Pearson.

Alloway, T. P., Gathercole, S. E., Kirkwood, H., & Elliott, J. (2009). The cognitive and behavioral characteristics of children with low working memory. *Child development*, *80*(2), 606-621.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders*, *4*.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub.

Anderson, P. (2002). Assessment and development of executive function (EF) during childhood. Child neuropsychology, 8(2), 71-82.

Archibald, S. J., & Kerns, K. A. (1999). Identification and description of new tests of executive functioning in children. *Child Neuropsychology*, *5*(2), 115-129.

Baddeley, A. (2000). The episodic buffer: a new component of working memory?. *Trends in cognitive sciences*, *4*(11), 417-423.

Barkley, R. A. (1997). Behavioral inhibition, sustained attention, and executive functions: constructing a unifying theory of ADHD. *Psychological bulletin*, *121*(1), 65.

Barkley, R. A. (2006). The relevance of the still lectures to attention-deficit/hyperactivity disorder: a commentary. *Journal of Attention Disorders*, *10*(2), 137-14

Barkley, R. A. (2011). The importance of emotion and ADHD. *Journal of ADHD and related disorders*, *1*(2), 5-37.

Barkley, R. A. (2015). Etiologies of ADHD. In R. A. Barkley (Eds.), Attention Deficit Hyperactivity Disorder: A Handbook for Diagnosis and Treatment (pp. 356-390). New York, NY: The Guilford Press.

Barkley, R. A., & Murphy, K. R. (2006). *Attention-deficit hyperactivity disorder: A clinical workbook*. Guilford Press.

Barkley, R. A., & Murphy, K. R. (2011). The nature of executive function (EF) deficits in daily life activities in adults with ADHD and their relationship to performance on EF tests. *Journal of Psychopathology and Behavioral Assessment*, *33*(2), 137-158.

Barkley, R. A., Fischer, M., Smallish, L., & Fletcher, K. (2006). Young adult outcome of hyperactive children: adaptive functioning in major life activities. *Journal of the American Academy of Child & Adolescent Psychiatry*, *45*(2), 192-202.

Barkley, R. A., Murphy, K. R., & Fischer, M. (2008). ADHD in adults: What the science says. New York, NY, US.

Barnard, H., Rao, R., Xu, Y., Froehlich, T., Epstein, J., Lanphear, B. P., & Yolton, K. (2018). Association of the Conners' Kiddie Continuous Performance Test (K-CPT) performance and parent-report measures of behavior and executive functioning. Journal of attention disorders, 22(11), 1056-1065.

Barrouillet, P., Gavens, N., Vergauwe, E., Gaillard, V., & Camos, V. (2009). Working memory span development: a time-based resource-sharing model account. *Developmental psychology*, *45*(2), 477.

Barry, T. D., Lyman, R. D. & Klinger, L. G. (2002) Academic underachievement and attention-deficit/hyperactivity disorder: the negative impact of symptom severity on school performance.

Berlin, L., & Bohlin, G. (2002). Response inhibition, hyperactivity, and conduct problems among preschool children. *Journal of Clinical Child and Adolescent Psychology*, *31*(2), 242-251.

Berwid, O. G., Curko Kera, E. A., Marks, D. J., Santra, A., Bender, H. A., & Halperin, J. M. (2005). Sustained attention and response inhibition in young children at risk for Attention Deficit/Hyperactivity Disorder. *Journal of Child Psychology and Psychiatry*, *46*(11), 1219-1229.

Biederman, J., Faraone, S., Milberger, S., Guite, J., Mick, E., Chen, L., ... & Spencer, T. (1996). A prospective 4-year follow-up study of attention-deficit hyperactivity and related disorders. Archives of general psychiatry, 53(5), 437-446.

Biederman, J., Petty, C. R., Fried, R., Doyle, A. E., Spencer, T., Seidman, L. J., ... & Faraone, S. V. (2007). Stability of executive function deficits into young adult years: A prospective longitudinal follow-up study of grown up males with ADHD. *Acta PsychiatricaScandinavica*, *116*(2), 129-136.

Biederman, J., Monuteaux, M.C., Doyle, A.E., Seidman, L.J., Wilens, T.E., Ferrero, F., Morgan, C.L. & Faraone, S.V. (2004) Impact of executive function deficits and attention-deficit/hyperactivity disorder (ADHD) on academic outcomes in children. Journal of Consulting and Clinical Psychology, 72, 757–766.

Biederman, J., Newcorn, J., & Sprich, S. (1991). Comorbidity of attention deficit hyperactivity disorder. *American journal of psychiatry*, *148*(5), 564-577.

Boonstra, A. M., Oosterlaan, J., Sergeant, J. A., & Buitelaar, J. K. (2005). Executive functioning in adult ADHD: a meta-analytic review. *Psychological medicine*, *35*(8), 1097-1108.

Breaux, R. P., Griffith, S. F., & Harvey, E. A. (2016). Preschool neuropsychological measures as predictors of later attention deficit hyperactivity disorder. Journal of abnormal child psychology, 44(8), 1455-1471.

Bresnahan, S. M., Anderson, J. W., & Barry, R. J. (1999). Age-related changes in quantitative EEG in attention-deficit/hyperactivity disorder. Biological Psychiatry, 46(12), 1690-1697.

Brocki, K. C., Eninger, L., Thorell, L. B., & Bohlin, G. (2010). Interrelations between executive function and symptoms of hyperactivity/impulsivity and inattention in preschoolers: A two year longitudinal study. Journal of abnormal child psychology, 38(2), 163-171.

Brooks, B. L., Sherman, E. M., & Strauss, E. (2009). NEPSY-II: A developmental neuropsychological assessment. *Child Neuropsychology*, *16*(1), 80-101.

Çak, H. T., Kültür, S. E. Ç., Gökler, B., Öktem, F., & Taşkıran, C. (2017). The behavior rating inventory of executive function and continuous performance test in preschoolers with attention deficit hyperactivity disorder. Psychiatry investigation, 14(3), 260.

Carlson, C. L., & Tamm, L. (2000). Responsiveness of children with attention deficit–hyperactivity disorder to reward and response cost: Differential impact on performance and motivation. *Journal of consulting and clinical psychology*, *68*(1), 73.

Castellanos, F. X., & Tannock, R. (2002). Neuroscience of attention-deficit/hyperactivity disorder: the search for endophenotypes. *Nature Reviews Neuroscience*, *3*(8), 617.

Castellanos, F. Xavier, Edmund JS Sonuga-Barke, Anouk Scheres, Adriana Di Martino, Christopher Hyde, and Judith R. Walters. "Varieties of attention-deficit/hyperactivity disorder-related intra-individual variability." *Biological psychiatry* 57, no. 11 (2005): 1416-1423.

Castle, L., Aubert, R. E., Verbrugge, R. R., Khalid, M., & Epstein, R. S. (2007). Trends in medication treatment for ADHD. *Journal of attention disorders*, *10*(4), 335-342.

Charach, A., Yeung, E., Climans, T., & Lillie, E. (2011). Childhood attention-deficit/hyperactivity disorder and future substance use disorders: comparative meta-analyses. *Journal of the American Academy of Child & Adolescent Psychiatry*, *50*(1), 9-21.

Collins, A., & Koechlin, E. (2012). Reasoning, learning, and creativity: frontal lobe function and human decision-making. *PLoS biology*, *10*(3), e1001293.

Conners, C. K. (1990). *Manual for Conner's Rating Scales*. Multi-Health Systems, Incorporated.

Conners, C. K. (2001). *Conners' Rating Scales--revised: CRS-R*. North Tonawanda, NJ: Multi-Health Systems.

Conners, C. K. (2008). Conners 3. MHS.

Conners, C. K., Sitarenios, G., Parker, J. D., & Epstein, J. N. (1998). The revised Conners' Parent Rating Scale (CPRS-R): factor structure, reliability, and criterion validity. *Journal of abnormal child psychology*, *26*(4), 257-268.

Conway, A. R., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic bulletin & review*, *12*(5), 769-786.

Cragg, L., & Nation, K. (2008). Go or no-go? Developmental improvements in the efficiency of response inhibition in mid-childhood. *Developmental Science*, *11*(6), 819-827.

Daley, D. (2010). Preschool-parent account of child symptoms (Pre-Pacs). *Unpublished manuscript*.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of verbal learning and verbal behavior*, *19*(4), 450-466.

Davis, H. L., & Pratt, C. (1995). The development of children's theory of mind: The working memory explanation. *Australian Journal of Psychology*, *47*(1), 25-31.

Dawson, G. (2008). Early behavioral intervention, brain plasticity, and the prevention of autism spectrum disorder. Development and Psychopathology, 20, 775–803.

De Luca, C. R., Wood, S. J., Anderson, V., Buchanan, J. A., Proffitt, T. M., Mahony, K., & Pantelis, C. (2003). Normative data from the CANTAB. I: development of executive function over the lifespan. *Journal of clinical and experimental neuropsychology*, *25*(2), 242-254.

Diamantopoulou, S., Rydell, A. M., Thorell, L. B., & Bohlin, G. (2007). Impact of executive functioning and symptoms of attention deficit hyperactivity disorder on children's peer relations and school performance. *Developmental neuropsychology*, *32*(1), 521-542.

Diamond, A. (2013). Executive functions. *Annual review of psychology*, *64*, 135-168.

Dickerson Mayes, S., Calhoun, L., & Crowell, E. W. (2001). Clinical validity and interpretation of the Gordon Diagnostic System in ADHD assessments. *Child Neuropsychology*, *7*(1), 32-41.

DiStefano, C., & Morgan, G. B. (2010). Evaluation of the BESS TRS-CA using the Rasch Rating Scale model. School Psychology Quarterly, 25(4), 202.

DuPaul, G. J. (2007). School-based interventions for students with attention deficit hyperactivity disorder: Current status and future directions. School Psychology Review, 36(2), 183

DuPaul, G. J., & Kern, L. (2011). *Young children with ADHD: Early identification and intervention*. American Psychological Association.

DuPaul, G. J., & Langberg, J. M. (2015). Educational impairments in children with ADHD.

DuPaul, G. J., Power, T. J., Anastopoulos, A. D., & Reid, R. (1998). *ADHD Rating Scale—IV: Checklists, norms, and clinical interpretation*. Guilford Press.

DuPaul, G. J., & Stoner, G. (2014). *ADHD in the schools: Assessment and intervention strategies*. Guilford Publications.

DuPaul, G. J., Volpe, R. J., Jitendra, A. K., Lutz, J. G., Lorah, K. S., & Gruber, R. (2004). Elementary school students with AD/HD: Predictors of academic achievement. Journal of School Psychology, 42(4), 285-301.

Eagle, D. M., Bari, A., & Robbins, T. W. (2008). The neuropsychopharmacology of action inhibition: cross-species translation of the stop-signal and go/no-go tasks. *Psychopharmacology*, *199*(3), 439-456.

Egger, H. L., & Angold, A. (2004). The Preschool Age Psychiatric Assessment (PAPA): A structured parent interview for diagnosing psychiatric disorders in preschool children. *Handbook of infant, toddler, and preschool mental health assessment*, *42*(5), 223-243.

Egger, H. L., Kondo, D., & Angold, A. (2006). The epidemiology and diagnostic issues in preschool attention-deficit/hyperactivity disorder: A review. *Infants & Young Children*, *19*(2), 109-122.

Elwood, R. W. (2001). MicroCog: assessment of cognitive functioning. *Neuropsychology review*, *11*(2), 89-100.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & psychophysics*, *16*(1), 143-149.

Frazier, T. W., Youngstrom, E. A., Glutting, J. J., & Watkins, M. W. (2007). ADHD and achievement: Meta-analysis of the child, adolescent, and adult literatures and a concomitant study with college students. *Journal of learning disabilities*, *40*(1), 49-65.

Forbes, G. B. (1998). Clinical utility of the test of variables of attention (TOVA) in the diagnosis of attention-deficit/hyperactivity disorder. *Journal of clinical psychology*, *54*(4), 461-476.

Frick, P. J., Kamphaus, R. W., Lahey, B. B., Loeber, R., Christ, M. A. G., Hart, E. L., & Tannenbaum, L. E. (1991). Academic underachievement and the disruptive behavior disorders. *Journal of consulting and clinical psychology*, *59*(2), 289.

Frye, D., Zelazo, P. D., & Palfai, T. (1995). Theory of mind and rule-based reasoning. *Cognitive development*, *10*(4), 483-527.

Gadow, K. D., & Nolan, E. E. (2002). Differences between preschool children with ODD, ADHD, and ODD+ ADHD symptoms. *Journal of Child Psychology and Psychiatry*, *43*(2), 191-201.

Gau, S. S. F., Shang, C. Y., Liu, S. K., Lin, C. H., Swanson, J. M., Liu, Y. C., & Tu, C. L. (2008). Psychometric properties of the Chinese version of the Swanson, Nolan, and Pelham, version IV scale–parent form. *International journal of methods in psychiatric research*, *17*(1), 35-44.

Gerstadt, C. L., Hong, Y. J., & Diamond, A. (1994). The relationship between cognition and action: performance of children 312–7 years old on a stroop-like day-night test. *Cognition*, *53*(2), 129-153.

Gioia, G., Isquith, P., & Guy, S. (2001). Assessment of executive functions in children with neurological impairment. In R. Simeonsson & S. Rosenthal (Eds.), Psychological and developmental assessment: Children with disabilities and chronic conditions (pp. 317–356). New York: The Guildford Press.

Graham-LoPresti, J. R., Williams, M. T., & Rosen, D. C. (2019). Culturally responsive assessment and diagnosis for clients of color.

Gremillion, M. L., Smith, T. E., & Martel, M. M. (2018). Verbal Working Memory as a Longitudinal Mechanism of Vocabulary Problems in Preschoolers with ADHD. Journal of Psychopathology and Behavioral Assessment, 40(1), 130-138.

Gualtieri, C. T., & Johnson, L. G. (2006). Reliability and validity of a computerized neurocognitive test battery, CNS Vital Signs. *Archives of Clinical Neuropsychology*, *21*(7), 623-643.

Gur, R. C., Erwin, R. J., Gur, R. E., Zwil, A. S., Heimberg, C., & Kraemer, H. C. (1992). Facial emotion discrimination: II. Behavioral findings in depression. *Psychiatry research*, *42*(3), 241-251.

Healey, D. M., Miller, C. J., Castelli, K. L., Marks, D. J., & Halperin, J. M. (2008). The impact of impairment criteria on rates of ADHD diagnoses in preschoolers. *Journal of Abnormal Child Psychology*, *36*(5), 771-778.

Henry, L., & MacLean, M. (2003). Relationships between working memory, expressive vocabulary and arithmetical reasoning in children with and without intellectual disabilities. *Educational and Child Psychology*, *20*(3), 51-63.

Hinshaw, S. P. (1992). Academic underachievement, attention deficits, and aggression: comorbidity and implications for intervention. *Journal of consulting and clinical psychology*, *60*(6), 893.

Hobson, C. W., Scott, S., & Rubia, K. (2011). Investigation of cool and hot executive function in ODD/CD independently of ADHD. *Journal of Child Psychology and Psychiatry*, *52*(10), 1035-1043.

Hollingsworth DE, McAuliffe SP, Knowlton BJ. Temporal allocation of visual attention in adult attention deficit hyperactivity disorder. Journal of Cognitive Neuroscience. 2001;13 (3):298–305.

Hommel, B. (2011). Attention and spatial stimulus coding in the Simon task: A rejoinder to van der Lubbe and Abrahamse (2010). *Acta psychologica*, *136*(2), 265-268.

Ikeda, Y., Hirata, S., Okuzumi, H., & Kokubun, M. (2010). Features of Stroop and reverse-Stroop interference: analysis by response modality and evaluation. *Perceptual and motor skills*, *110*(2), 654-660.

Jacobson, L. A., Schneider, H., & Mahone, E. M. (2017). Preschool Inhibitory Control Predicts ADHD Group Status and Inhibitory Weakness in School. Archives of Clinical Neuropsychology, 33(8), 1006-1014.

Jang, J., Matson, J. L., Williams, L. W., Tureck, K., Goldin, R. L., & Cervantes, P. E. (2013). Rates of comorbid symptoms in children with ASD, ADHD, and comorbid ASD and ADHD. *Research in developmental disabilities*, *34*(8), 2369-2378.

Kaufman, A. S., & Kaufman, N. L. (1983). *K-ABC: Kaufman assessment battery for children: Interpretive manual*. American Guidance Service.

Kaufman, J., Birmaher, B., Brent, D., Rao, U. M. A., Flynn, C., Moreci, P., ... & Ryan, N. (1997). Schedule for affective disorders and schizophrenia for school-age children-present and lifetime version (K-SADS-PL): initial reliability and validity data. *Journal of the American Academy of Child & Adolescent Psychiatry*, *36*(7), 980-988.

Kaufman, J., Birmaher, B., Brent, D., Rao, U. M. A., Flynn, C., Moreci, P., ... & Ryan, N. (1997). Schedule for affective disorders and schizophrenia for school-age children-present and lifetime version (K-SADS-PL): initial reliability and validity data. *Journal of the American Academy of Child & Adolescent Psychiatry*, *36*(7), 980-988.

Kaufmann, L. & Nuerk, H. C. (2008) Basic number processing deficits in ADHD: a broad examination of elementary and complex number processing skills in 9- to 12-year old children with ADHD-C. Developmental Science, 11, 692–699.

Kelly, T. (2000). The development of executive function in school-aged children. Clinical Neuropsychological Assessment, 1, 38–55.

Kochanska, G., Coy, K. C., & Murray, K. T. (2001). The development of self-regulation in the first four years of life. *Child development*, *72*(4), 1091-1111.

Koglin, U., Barquero, B., Mayer, H., Scheithauer, H., & Petermann, F. (2007). Deutsche Version des Strengths and Difficulties Questionnaire (SDQ-Deu). *Diagnostica*, *53*(4), 175-183.

Korkman, M., & Pesonen, A. E. (1994). A comparison of neuropsychological test profiles of children with attention deficit—hyperactivity disorder and/or learning disorder. *Journal of Learning Disabilities*, *27*(6), 383-392.

Kuntsi, J., Oosterlaan, J., & Stevenson, J. (2001). Psychological mechanisms in hyperactivity: I response inhibition deficit, working memory impairment, delay aversion, or something else?. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, *42*(2), 199-210.

Lahey, B. B., Pelham, W. E., Loney, J., Kipp, H., Ehrhardt, A., Lee, S. S., ... & Massetti, G. (2004). Three-year predictive validity of DSM-IV attention deficit hyperactivity disorder in children diagnosed at 4–6 years of age. *American Journal of Psychiatry*, *161*(11), 2014-2020.

Langenecker, S. A., Bieliauskas, L. A., Rapport, L. J., Zubieta, J. K., Wilde, E. A., & Berent, S. (2005). Face emotion perception and executive functioning deficits in depression. *Journal of clinical and experimental neuropsychology*, *27*(3), 320-333.

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

Levy, F., Hay, D. A., Bennett, K. S., & McStephen, M. (2005). Gender differences in ADHD subtype comorbidity. *Journal of the American Academy of Child & Adolescent Psychiatry*, *44*(4), 368-376.

Li, C. S. R., Chen, S. H., Lin, W. H., & Yang, Y. Y. (2005). Attentional blink in adolescents with varying levels of impulsivity. *Journal of Psychiatric Research*, *39*(2), 197-205.

Lonigan, C. J., Spiegel, J. A., Goodrich, J. M., Morris, B. M., Osborne, C. M., Lerner, M. D., & Phillips, B. M. (2017). Does preschool self-regulation predict later behavior problems in general or specific problem behaviors?. Journal of abnormal child psychology, 45(8), 1491-1502.

Lu, C. H., & Proctor, R. W. (1995). The influence of irrelevant location information on performance: A review of the Simon and spatial Stroop effects. *Psychonomic bulletin & review*, *2*(2), 174-207.

Lucas, C. P., Fisher, P., & Luby IV, J. L. (2008). Young child DISC-IV Diagnostic Interview Schedule for Children. *New York: Division of Child Psychiatry, Colombia University*.

Luciana, M., & Nelson, C. A. (2002). Assessment of neuropsychological function through use of the Cambridge Neuropsychological Testing Automated Battery:

performance in 4-to 12-year-old children. *Developmental neuropsychology*, *22*(3), 595-624.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: an integrative review. *Psychological bulletin*, *109*(2), 163.

Mahone, E. M., Pillion, J. P., & Hiemenz, J. R. (2001). Initial development of an auditory continuous performance test for preschoolers. *Journal of attention disorders*, *5*(2), 93-106.

Massetti, G. M., Lahey, B. B., Pelham, W. E., Loney, J., Ehrhardt, A., Lee, S. S., & Kipp, H. (2008). Academic achievement over 8 years among children who met modified criteria for attention-deficit/hyperactivity disorder at 4–6 years of age. *Journal of abnormal child psychology*, *36*(3), 399-410.

Maqableh, M. (2015). The acceptance and use of computer based assessment in higher education. Journal of Software Engineering and Applications, 8(10), 557.

Mayes, S. D., Calhoun, S. L., & Crowell, E. W. (2000). Learning disabilities and ADHD: Overlapping spectrum disorders. *Journal of learning disabilities*, *33*(5), 417-424.

McCarthy, D. A. (1970). *McCarthy scales of children's abilities*. Psychological Corporation.

McGee, R., Williams, S., & Feehan, M. (1992). Attention Deficit Disorder and age of onset of problem behaviors. Journal of Abnormal Child Psychology, 20(5), 487–502

McInerney, R. J., Hrabok, M., & Kerns, K. A. (2005). The children’s size-ordering task: a new measure of nonverbal working memory. *Journal of Clinical and Experimental Neuropsychology*, *27*(6), 735-745.

Mendeley. (2009). Getting Started with Mendeley. Mendeley Desktop. London: Mendeley Ltd. Retrieved from http://www. mendeley.com. meta-analytic review. *Neurotherapeutics*, *9*(3), 490-499.

Milich, R., Balentine, A. C., & Lynam, D. R. (2001). ADHD combined type and ADHD predominantly inattentive type are distinct and unrelated disorders. *Clinical psychology: science and practice*, *8*(4), 463-488.

Miranda, A., Colomer, C., Mercader, J., Fernández, M. I., & Presentación, M. J. (2015). Performance-based tests versus behavioral ratings in the assessment of executive functioning in preschoolers: associations with ADHD symptoms and reading achievement. *Frontiers in psychology*, *6*, 545.

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current directions in psychological science*, *21*(1), 8-14.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive psychology*, *41*(1), 49-100.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of internal medicine*, *151*(4), 264-269.

Motamedi, M., Bierman, K., & Huang-Pollock, C. L. (2016). Rejection reactivity, executive function skills, and social adjustment problems of inattentive and hyperactive kindergarteners. Social Development, 25(2), 322-339.

Morris, N., & Jones, D. M. (1990). Memory updating in working memory: The role of the central executive. *British journal of psychology*, *81*(2), 111-121.

Mullane, J. C., Corkum, P. V., Klein, R. M., & McLaughlin, E. (2009). Interference control in children with and without ADHD: a systematic review of Flanker and Simon task performance. *Child Neuropsychology*, *15*(4), 321-342.

Naglieri, J. A., Drasgow, F., Schmit, M., Handler, L., Prifitera, A., Margolis, A., & Velasquez, R. (2004). Psychological testing on the Internet: new problems, old issues. *American Psychologist*, *59*(3), 150.

Nigg, J. T., Stavro, G., Ettenhofer, M., Hambrick, D. Z., Miller, T., & Henderson, J. M. (2005). Executive functions and ADHD in adults: evidence for selective effects on ADHD symptom domains. *Journal of Abnormal Psychology*, *114*(4), 706.

Owen, A. M., McMillan, K. M., Laird, A. R., & Bullmore, E. (2005). N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. *Human brain mapping*, *25*(1), 46-59.

Pauli-Pott, U., & Becker, K. (2011). Neuropsychological basic deficits in preschoolers at risk for ADHD: A meta-analysis. *Clinical Psychology Review*, *31*(4), 626-637.

Pauli-Pott, U., Dalir, S., Mingebach, T., Roller, A., & Becker, K. (2014). Attention deficit/hyperactivity and comorbid symptoms in preschoolers: Differences between subgroups in neuropsychological basic deficits. Child Neuropsychology, 20(2), 230-244.

Pelham, Jr, W. E., Fabiano, G. A., & Massetti, G. M. (2005). Evidence-based assessment of attention deficit hyperactivity disorder in children and adolescents. *Journal of clinical child and adolescent psychology*, *34*(3), 449-476.

Pliszka, S. R. (1998). Comorbidity of attention-deficit/hyperactivity disorder with psychiatric disorder: an overview. *The Journal of clinical psychiatry*, *59*, 50-58.

Polanczyk, G., De Lima, M. S., Horta, B. L., Biederman, J., & Rohde, L. A. (2007). The worldwide prevalence of ADHD: a systematic review and metaregressionanalysis. *American journal of psychiatry*, *164*(6), 942-948.

Posner, M. I., & DiGirolamo, G. J. (1998). 1 8 Executive Attention: Conflict, Target Detection, and Cognitive Control.

Raggi, V. L., & Chronis, A. M. (2006). Interventions to address the academic impairment of children and adolescents with ADHD. Clinical Child and Family Psychology Review, 9(2), 85-111.

Rajendran, K., O'neill, S., Marks, D. J., & Halperin, J. M. (2015). Latent profile analysis of neuropsychological measures to determine preschoolers' risk for ADHD. Journal of Child Psychology and Psychiatry, 56(9), 958-965.

Rapport, M. D., Orban, S. A., Kofler, M. J., & Friedman, L. M. (2013). Do programs designed to train working memory, other executive functions, and attention benefit children with ADHD? A meta-analytic review of cognitive, academic, and behavioral outcomes. *Clinical psychology review*, *33*(8), 1237-1252.

Rapport, M. D., Scanlan, S. W., & Denney, C. B. (1999). Attention-deficit/hyperactivity disorder and scholastic achievement: A model of dual developmental pathways. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, *40*(8), 1169-1183.

Reich, W., Welner, Z., & Herjanic, B. (1997). Diagnostic interview for children-IV (DICA-IV). *Toronto, Ontario, Canada: Multi-Health Systems Incorporated*.

Rennie, B., Beebe-Frankenberger, M., & Swanson, H. L. (2014). A longitudinal study of neuropsychological functioning and academic achievement in children with and without signs of attention-deficit/hyperactivity disorder. Journal of clinical and experimental neuropsychology, 36(6), 621-635.

Resnick, R. J. (2005). Attention deficit hyperactivity disorder in teens and adults: They don't all outgrow it. Journal of clinical psychology, 61(5), 529-533.

Rezazadeh, S. M., Wilding, J., & Cornish, K. (2011). The relationship between measures of cognitive attention and behavioral ratings of attention in typically developing children. Child Neuropsychology, 17(2), 197-208.

Riccio, C. A., Cohen, M. J., Hynd, G. W., & Keith, R. W. (1996). Validity of the auditory continuous performance test in differentiating central processing auditory disorders with and without ADHD. *Journal of Learning Disabilities*, *29*(5), 561-566.

Robb, J. A., Sibley, M. H., Pelham, W. E., Foster, E. M., Molina, B. S., Gnagy, E. M., & Kuriyan, A. B. (2011). The estimated annual cost of ADHD to the US education system. School mental health, 3(3), 169-177.

Roberts, W., Milich, R., & Barkley, R. A. (2015). Primary symptoms, diagnostic criteria, subtyping, and prevalence of ADHD. *Attention-deficit hyperactivity disorder*, *4*, 50-80.

Sanchez, R. J., Crismon, M. L., Barner, J. C., Bettinger, T., & Wilson, J. P. (2005). Assessment of adherence measures with different stimulants among children and adolescents. *Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy*, *25*(7), 909-917.

Sasser, T. R., Beekman, C. R., & Bierman, K. L. (2015). Preschool executive functions, single-parent status, and school quality predict diverging trajectories of classroom inattention in elementary school. Development and psychopathology, 27(3), 681-693.

Schachar, R., Mota, V. L., Logan, G. D., Tannock, R., & Klim, P. (2000). Confirmation of an inhibitory control deficit in attention-deficit/hyperactivity disorder. *Journal of abnormal child psychology*, *28*(3), 227-235.

Schatz, D. B., & Rostain, A. L. (2006). ADHD with comorbid anxiety: a review of the current literature. *Journal of Attention disorders*, *10*(2), 141-149.

Scheeringa, M. S., & Haslett, N. (2010). The reliability and criterion validity of the Diagnostic Infant and Preschool Assessment: a new diagnostic instrument for young children. *Child Psychiatry & Human Development*, *41*(3), 299-312.

Schoemaker, K., Mulder, H., Deković, M., & Matthys, W. (2013). Executive functions in preschool children with externalizing behavior problems: A meta-analysis. *Journal of abnormal child psychology*, *41*(3), 457-471.

Seidman, L. J., Biederman, J., Faraone, S. V., Milberger, S., Norman, D., Seiverd, K., ... & Kiely, K. (1995). Effects of family history and comorbidity on the neuropsychological performance of children with ADHD: Preliminary findings. *Journal of the American Academy of Child & Adolescent Psychiatry*, *34*(8), 1015-1024.

Sergeant, J. A., Geurts, H., Huijbregts, S., Scheres, A., & Oosterlaan, J. (2003). The top and the bottom of ADHD: a neuropsychological perspective. *Neuroscience & Biobehavioral Reviews*, *27*(7), 583-592.

Simonoff, E., Pickles, A., Wood, N., Gringras, P. & Chadwick, O. (2007) ADHD symptoms in children with mild intellectual disability. Journal of the American Academy of Child and Adolescent Psychiatry, 46, 591–600.

Simpson, A., & Riggs, K. J. (2005). Inhibitory and working memory demands of the day–night task in children. *British Journal of Developmental Psychology*, *23*(3), 471-486.

Sjöwall, D., Bohlin, G., Rydell, A. M., & Thorell, L. B. (2017). Neuropsychological deficits in preschool as predictors of ADHD symptoms and academic achievement in late adolescence. *Child Neuropsychology*, *23*(1), 111-128.

Sjöwall, D., Backman, A., & Thorell, L. B. (2015). Neuropsychological heterogeneity in preschool ADHD: Investigating the interplay between cognitive, affective and motivation-based forms of regulation. Journal of abnormal child psychology, 43(4), 669-680.

Sjöwall, D., Bohlin, G., Rydell, A. M., & Thorell, L. B. (2017). Neuropsychological deficits in preschool as predictors of ADHD symptoms and academic achievement in late adolescence. Child Neuropsychology, 23(1), 111-128.

Skogan, A. H., Zeiner, P., Egeland, J., Rohrer-Baumgartner, N., Urnes, A. G., Reichborn-Kjennerud, T., & Aase, H. (2014). Inhibition and working memory in young preschool children with symptoms of ADHD and/or oppositional-defiant disorder. *Child Neuropsychology*, *20*(5), 607-624.

Skogli, E. W., Teicher, M. H., Andersen, P. N., Hovik, K. T., & Øie, M. (2013). ADHD in girls and boys–gender differences in co-existing symptoms and executive function measures. *BMC psychiatry*, *13*(1), 298.

Sonuga-Barke, E. J. S. (2002) Psychological heterogeneity in AD/HD – a dual pathway model of behaviour and cognition. Behavioural Brain Research, 130, 29–36.

Sonuga-Barke, E. J., & Halperin, J. M. (2010). Developmental phenotypes and causal pathways in attention deficit/hyperactivity disorder: potential targets for early intervention?. *Journal of Child Psychology and Psychiatry*, *51*(4), 368-389.

Sonuga-Barke, E. J. S., Houlberg, K., & Hall, M. (1994). When is “impulsiveness” not impulsive? The case of hyperactive children's cognitive style. *Journal of Child Psychology and Psychiatry*, *35*(7), 1247-1253.

Spencer, T. J. (2006). ADHD and comorbidity in childhood. *The Journal of clinical psychiatry*, *67*, 27-31.

Sprafkin, J., & Gadow, K. (1997). Early Childhood Inventory 4 (ECI-4). *Stony Brook, NY: Checkmate Plus*.

Stefanatos, G. A., & Baron, I. S. (2007). Attention-deficit/hyperactivity disorder: A neuropsychological perspective towards DSM-V. *Neuropsychology review*, *17*(1), 5-38.

Stuss, D. T., Benson, D. F., Clermont, R., Della Malva, C. L., Kaplan, E. F., & Weir, W. S. (1986). Language functioning after bilateral prefrontal leukotomy. *Brain and language*, *28*(1), 66-70.

Stuss, D. T., Levine, B., Alexander, M. P., Hong, J., Palumbo, C., Hamer, L., ... & Izukawa, D. (2000). Wisconsin Card Sorting Test performance in patients with focal frontal and posterior brain damage: effects of lesion location and test structure on separable cognitive processes. *Neuropsychologia*, *38*(4), 388-402.

Tinius, T. P. (2003). The integrated visual and auditory continuous performance test as a neuropsychological measure. *Archives of clinical Neuropsychology*, *18*(5), 439-454.

Theeuwes, J. (1991). Cross-dimensional perceptual selectivity. *Perception & psychophysics*, *50*(2), 184-193.

Thompson, A. L., Molina, B. S., Pelham Jr, W., & Gnagy, E. M. (2007). Risky driving in adolescents and young adults with childhood ADHD. *Journal of pediatric psychology*, *32*(7), 745-759.

Thorell, L. B. (2007) Do delay aversion and executive function deficits make distinct contributions to the functional impact of ADHD symptoms? A study of early academic skill deficits. Journal of Child Psychology and Psychiatry, 48, 1061–1070.

Van Lieshout, M., Luman, M., Buitelaar, J., Rommelse, N. N. J., & Oosterlaan, J. (2013). Does neurocognitive functioning predict future or persistence of ADHD? A systematic review. *Clinical psychology review*, *33*(4), 539-560.

Votruba, K. L., & Langenecker, S. A. (2013). Factor structure, construct validity, and age-and education-based normative data for the Parametric Go/No-Go Test. *Journal of clinical and experimental neuropsychology*, *35*(2), 132-146.

Van Mourik, R., Oosterlaan, J., & Sergeant, J. A. (2005). The Stroop revisited: A meta-analysis of interference control in AD/HD. *Journal of Child Psychology and Psychiatry*, *46*(2), 150-165.

Verbruggen, F., & Logan, G. D. (2008). Response inhibition in the stop-signal paradigm. *Trends in cognitive sciences*, *12*(11), 418-424.

Wåhlstedt, C., Thorell, L. B., & Bohlin, G. (2009). Heterogeneity in ADHD: Neuropsychological pathways, comorbidity and symptom domains. *Journal of abnormal child psychology*, *37*(4), 551-564.

Watkins, M. W., Glutting, J. J., & Lei, P. W. (2007). Validity of the Full-Scale IQ when there is significant variability among WISC-III and WISC-IV factor scores. *Applied Neuropsychology*, *14*(1), 13-20.

Wechsler, D. (1991). *WISC-III: Wechsler intelligence scale for children: Manual*. Psychological Corporation.

Welsh, M., Pennington, B., & Groisser, D. (1991). A normative-developmental study of executive function: A window on prefrontal function in children. Developmental Neuropsychology, 7, 131–149.

Weidacker, K., Whiteford, S., Boy, F., & Johnston, S. J. (2017). Response inhibition in the parametric go/no-go task and its relation to impulsivity and subclinical psychopathy. *The Quarterly Journal of Experimental Psychology*, *70*(3), 473-487.

Weyandt, L. L., & Gudmundsdottir, B. G. (2015). Developmental and neuropsychological deficits in children with ADHD. *Attention-deficit hyperactivity disorder, a handbook for diagnosis and treatment, 4th edn. The Guilford Press, New York*, 116-139.

Wilens, T. E., Biederman, J., Mick, E., Faraone, S. V., & Spencer, T. (1997). Attention deficit hyperactivity disorder (ADHD) is associated with early onset substance use disorders. *The Journal of nervous and mental disease*, *185*(8), 475-482.

Willcutt, E. G., Doyle, A. E., Nigg, J. T., Faraone, S. V., & Pennington, B. F. (2005). Validity of the executive function theory of attention-deficit/hyperactivity disorder: a meta-analytic review. *Biological psychiatry*, *57*(11), 1336-1346.

Wood, A. C., Asherson, P., Rijsdijk, F., & Kuntsi, J. (2009). Is overactivity a core feature in ADHD? Familial and receiver operating characteristic curve analysis of mechanically assessed activity level. *Journal of the American Academy of Child & Adolescent Psychiatry*, *48*(10), 1023-1030.

Zhang, H. F., Shuai, L., Zhang, J. S., Wang, Y. F., Lu, T. F., Tan, X., ... & Shen, L. X. (2018). Neuropsychological Profile Related with Executive Function of Chinese Preschoolers with Attention-Deficit/Hyperactivity Disorder: Neuropsychological Measures and Behavior Rating Scale of Executive Function-Preschool Version. Chinese medical journal, 131(6), 648.

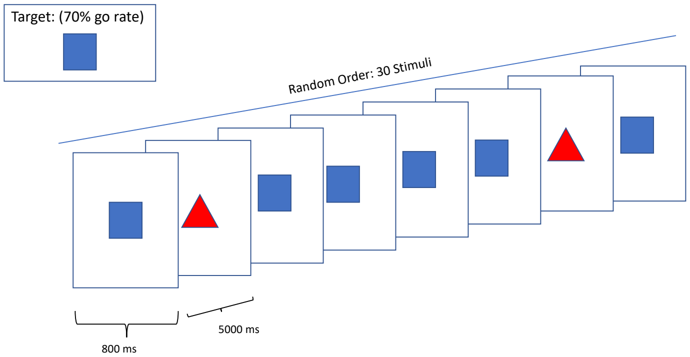
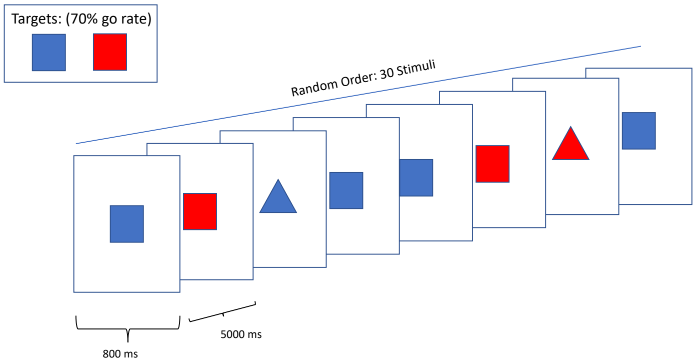
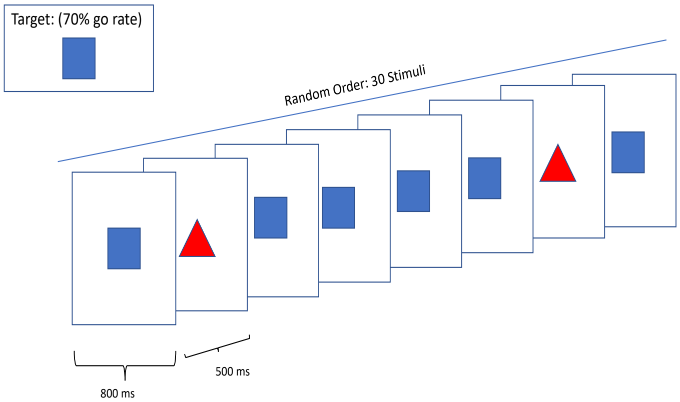
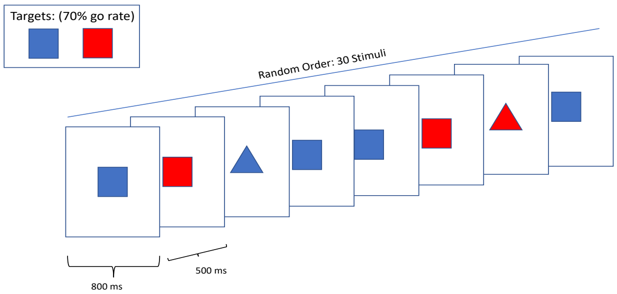
Zuvekas, S. H., & Vitiello, B. (2012). Stimulant medication use in children: a 12-year perspective. American Journal of Psychiatry, 169(2), 160-166.

Appendix A: Tables and Figures

Table A.1 *Developmental Models of EF*

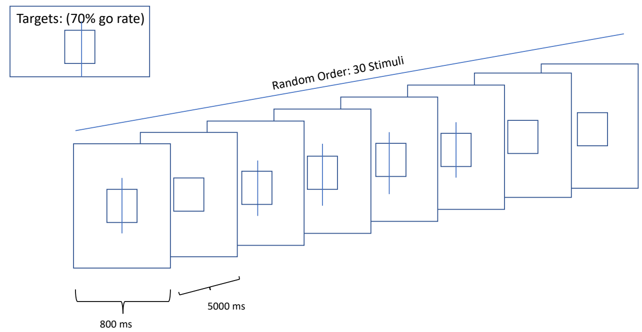
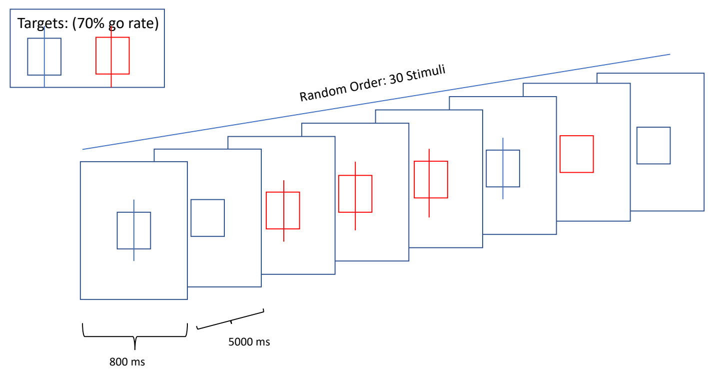
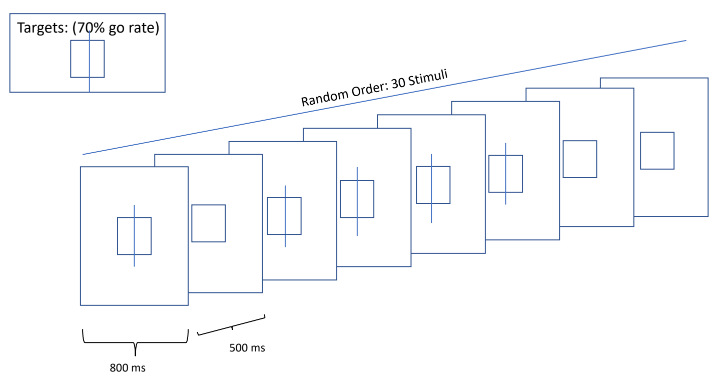
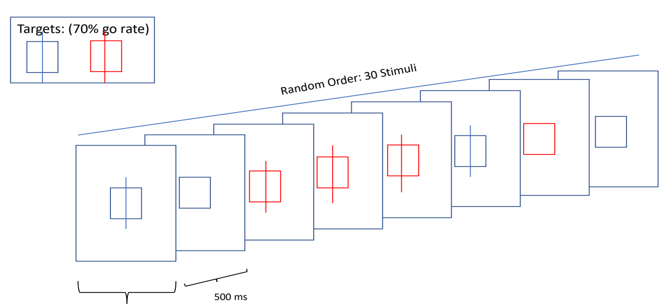
Author	Construct	Definition
Welsh and colleagues (1991)	(1) Speeded Responding	(1)The ability to react quickly
	(2) Set Maintenance	(2) The skill of controlling impulses
	(3)Planning	(3) The ability to prepare actions towards a future goal
Levin and colleagues (1991)	(1)Semantic Reasoning	(1) "Capacity to utilize semantic features when seeking information" (p. 391)
	(2)Freedom from perseveration	(2) Ability to inhibit prepotent response
	(3)Planning	(3) Ability to Formulate a strategy
Anderson (2002)	(1) Attentional Control	(1) the ability to attend to stimuli and inhibit prepotent responses; the capacity to maintain focused attention
	(2) Information Processing	(2) fluency, efficiency, and speed of response
	(3) Cognitive Flexibility	(3) "...the ability to shift between response sets, learn from mistakes, devise alternative strategies, divide attention, and process multiple sources of information concurrently" (p. 74). Working memory is contained here
	(4) Goal Setting	(4) the ability to develop plans in advance and begin undertakings in an effective and thoughtful way.
Miyake (2000)	(1) Shifting	(1) Switching between tasks or mental sets
	(2) Updating	(2) Monitoring and encoding information for the task at hand; related to working memory
	(3) Inhibition	(3) Purposefully restrain dominant responses
Garon and colleagues (2008)	(1)Working Memory	(1) the ability to hold/update/manipulate information in mind
	(2) Inhibition	(2)Withholding an automatic response
	(3) Shifting	(3) Switching attention between mental sets and/or tasks

Table A.2 *Level 1 Go/No-Go Task- High Discrimination*

Block	Parameters	Graphic
Block 1 *Difficulty 1*	Long Inter-stimulus Interval and Low Number of Distractors	
Block 3 *Difficulty 2*	Long Inter-stimulus Interval and High Number of Distractors	
Block 2 *Difficulty 3*	Short Inter-stimulus Interval and Low Number of Distractors	
Block 4 *Difficulty 4*	Short inter-stimulus interval and High Number of Distractors	

Note: Blocks were reordered from theoretical difficulty hierarchy to reduce load on cognitive flexibility

Table A.3 *Level 2 Go/No-Go Task- Low Discrimination*

Block	Parameters	Graphic
Block 5 *Difficulty 5*	Long Inter-stimulus Interval and Low Number of Distractors	
Block 7 *Difficulty 6*	Long Inter-stimulus Interval and High Number of Distractors	
Block 6 *Difficulty 7*	Short Inter-stimulus Interval and Low Number of Distractors	
Block 8 *Difficulty 8*	Short inter-stimulus interval and High Number of Distractors	

Note: Blocks were reordered from theoretical difficulty hierarchy to reduce load on cognitive flexibility

Table A.4. Demographic Characteristics of the Sample

	n	*Age in Years M (SD)*	*Females (%)*	*Males (%)*	*White (%)*	*Black (%)*	*Asian (%)*	*Hawaiian/ Pacific Islander (%)*	*American Indian/ Alaskan Native (%)*	*Two or More Races (%)*
Early Childhood	13 (10.92%)	5.60 (1.08)	7 (53.85%)	6 (46.15%)	12 (92.31%)	--	--	--	--	1 (7.69%)
Middle Childhood	10 (8.40%)	7.90 (0.61)	4 (40.00%)	6 (60.00%)	10 (100.00%)	--	--	--	--	--
Late Childhood	16 (13.46%)	10.60 (0.91)	8 (50.00%)	8 (50.00%)	14 (87.50%)	--	--	--	1 (6.25%)	1 (6.25%)
Early Adolescence	9 (7.56%)	13.70 (1.42)	4 (44.44%)	5 (55.56%)	8 (88.89%)	--	--	--	1 (11.11%)	--
Late Adolescence	19 (15.97%)	18.20 (1.36)	12 (63.16%)	7 (36.84%)	14 (73.68%)	--	3 (15.79%)	--	1 (5.26%)	1 (5.26%)
Young Adults	52 (43.70%)	25.70 (4.00)	34 (65.38%)	18 (34.62%)	37 (71.15%)	5 (9.62%)	1 (1.92%)	1 (1.92%)	3 (5.77%)	5 (9.62%)
Total Sample	119	17.4 (8.56)	60 (59.00%)	50 (42.00%)	95 79.80%	5 (4.20%)	4 (3.40%).	1 (0.80%)	6 (5.00%)	8 (6.70%)

‘

Table A.5 Average *Total Commission Score*

	Females M (SD)	*Males M (SD)*	*White M (SD)*	*Black M (SD)*	*Asian M (SD)*	*Hawaiian/ Pacific Islander M (SD)*	*American Indian/ Alaskan Native M (SD)*	*Two or More Race M (SD)*
Early Childhood	38.00 (19.90)	42.50 (8.64)	42.60 (12.90)	--	--	--	--	10.00*
Middle Childhood	56.80 (7.37)	50.80 (9.04)	53.20 (8.53)	--	--	--	--	--
Late Childhood	59.40 (8.09)	59.40 (13.40)	56.40 (10.50)	--	--	--	70.00*	70.00*
Early Adolescence	61.30 (11.40)	64.50 (8.35)	62.30 (9.57)	--	--	--	70.00*	--
Late Adolescence	60.80 (22.10)	68.30 (4.07)	61.20 (20.40)	--	69.30 (1.53)	69.30 (1.53)	72.00*	71.00*
Young Adults	64.10 (14.60)	68.90 (3.19)	68.90 (20.40)	57.40 (21.90)	68.00*	15.00*	69.00 (1.73)	59.20 (19.90)

*n = 1; *SD* not reported

Table A.6. Average *Block Commission Score*

Block	*1* *M* *(SD)*	*2* *M* *(SD)*	*3* *M* *(SD)*	*4* *M* *(SD)*	*5* *M* *(SD)*	*6* *M* *(SD)*	*7* *M* *(SD)*	*8* *M* *(SD)*
Early Childhood	6.46 (1.39)	5.92 (2.02)	5.62 (2.60)	5.31 (2.29)	4.62 (2.40)	3.85 (2.27)	4.69 (2.75)	3.62 (3.01)
Middle Childhood	7.80 (1.03)	7.00 (1.49)	7.30 (0.95)	6.70 (1.64)	7.10 (1.66)	5.40 (2.80)	6.20 (1.32)	5.70 (2.21)
Late Childhood	8.19 (1.76)	7.63 (1.31)	7.75 (1.18)	7.31 (1.14)	7.88 (1.09)	6.69 (2.06)	6.75 (2.29)	5.94 (2.57)
Early Adolescence	8.22 (1.39)	8.11 (1.17)	8.33 (1.00)	7.56 (1.67)	8.00 (2.65)	6.89 (2.89)	7.89 (1.05)	8.11 (0.93)
Late Adolescence	8.47 (1.43)	7.84 (2.79)	8.05 (2.12)	7.79 (2.78)	8.21 (2.07)	7.74 (2.66)	7.68 (2.65)	7.79 (2.68)
Young Adults	8.54 (1.36)	8.75 (0.68)	8.25 (1.70)	8.23 (1.55)	7.85 (2.12)	8.15 (2.11)	7.94 (2.10)	8.08 (2.31)

Table A.7 Two-way ANOVA - Total Commission Score

Factor	*SS*	*df*	*F*	*p*
Age group	7876.00	5	9.171	<.001
Gender	79.00	1	79.00	.50
Age group x Gender	445.90	5	89.20	.76
Residuals	18377.40	107	171.80	

Table A.8 Pairwise Comparisons- *Total Commission Score*

		Middle Childhood	Late Childhood	Early Adolescence	Late Adolescence	Early Adulthood
Early Childhood	Mean Difference	-13.54	-17.87	-22.68	-24.31	-26.28
	t	-2.43	-3.65	3.97	-5.07	-6.39
	p	0.157	0.005	0.002	<.001	<.001
Middle Childhood	Mean Difference		-4.33	-9.13	-10.77	-12.74
	t		0.81	1.50	2.05	-2.74
	p		0.965	0.667	0.332	0.075
Late Childhood	Mean Difference			-4.80	-6.43	-8.41
	t			0.88	1.42	-2.21
	p			0.952	0.713	0.239
Early Adolescence	Mean Difference				-1.63	-3.61
	t				-0.30	-0.75
	p				1.000	0.975
Late Adolescence	Mean Difference					-1.97
	t					-0.54
	p					0.994

Table A.9. Average *Total Omission Score*

Block	*1* *M* *(SD)*	*2* *M* *(SD)*	*3* *M* *(SD)*	*4* *M* *(SD)*	*5* *M* *(SD)*	*6* *M* *(SD)*	*7* *M* *(SD)*	*8* *M* *(SD)*
Early Childhood	16.50	17.80	14.8	16.50	15.40	16.30	15.50	13.50
	(3.87)	(2.89)	(5.71)	(5.78)	(5.94)	(5.74)	(5.77)	(6.72)
Middle Childhood	17.40	19.30	18.30	19.00	16.20	18.10	19.50	16.10
	(2.63)	(1.64)	(2.41)	(1.94)	(4.21)	(2.92)	(1.27)	(4.84)
Late Childhood	18.30	18.60	19.20	18.70	19.20	16.60	19.30	17.70
	(2.49)	(2.85)	(2.66)	(2.52)	(3.12)	(5.40)	(2.55)	(3.91)
Early Adolescence	20.30	20.60	20.70	20.10	18.30	18.70	20.40	19.80
	(0.87)	(0.73)	(0.50)	(1.36)	(6.63)	(7.00)	(1.33)	(0.67)
Late Adolescence	19.30	18.60	19.50	19.50	19.60	19.30	19.70	18.80
	(3.38)	(6.57)	(4.79)	(1.09)	(4.79)	(4.90)	(4.79)	(4.62)
Young Adults	19.80	20.40	19.80	19.90	19.70	19.80	19.30	18.60
	(1.92)	(2.28)	(3.57)	(3.56)	(4.14)	(4.08)	(5.04)	(5.10)

Table A.10. Average *Total PCT*

Block	*1* *M* *(SD)*	*2* *M* *(SD)*	*3* *M* *(SD)*	*4* *M* *(SD)*	*5* *M* *(SD)*	*6* *M* *(SD)*	*7* *M* *(SD)*	*8* *M* *(SD)*
Early Childhood	0.77	0.79	0.68	0.73	0.67	0.67	0.66	0.57
	(0.15)	(0.14)	(0.25)	(0.26)	(0.26)	(0.23)	(0.24)	(0.25)
Middle Childhood	0.84	0.88	0.85	0.86	0.77	0.78	0.84	0.73
	(0.10)	(0.09)	(0.10)	(0.10)	(0.14)	(0.15)	(0.06)	(0.19)
Late Childhood	0.88	0.88	0.90	0.87	0.90	0.78	0.85	0.79
	(0.11)	(0.12)	(0.11)	(0.10)	(0.11)	(0.18)	(0.12)	(0.14)
Early Adolescence	0.95	0.96	0.97	0.92	0.88	0.85	0.93	0.93
	(0.06)	(0.05)	(0.04)	(0.05)	(0.31)	(0.32)	(0.35)	(0.05)
	0.92	0.88	0.92	0.93	0.93	0.91	0.90	0.90
Late Adolescence	(0.19)	(0.31)	(0.22)	(0.23)	(0.23)	(0.23)	(0.22)	(0.22)
	0.95	0.97	0.93	0.94	0.92	0.94	0.88	0.89
Young Adults	(0.09)	(0.08)	(0.17)	(0.16)	(0.20)	(0.19)	(0.23)	(0.24)

Table A.11. Average *Total RT*

Block	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
	M	*M*	*M*	*M*	*M*	*M*	*M*	*M*
	(SD)	*(SD)*	*(SD)*	*(SD)*	*(SD)*	*(SD)*	*(SD)*	*(SD)*
Early Childhood	1.12	0.63	1.24	0.64	1.12	0.64	1.14	0.68
	(0.41)	(0.22)	(0.47)	(0.19)	(0.43)	(0.18)	(0.36)	(0.20)
Middle Childhood	0.94	0.59	1.10	0.63	1.07	0.57	1.13	0.66
	(0.27)	(0.21)	(0.32)	(0.18)	(0.41)	(0.12)	(0.46)	(0.15)
Late Childhood	0.99	0.64	0.93	0.62	0.94	0.69	0.96	0.57
	(0.41)	(0.22)	(0.26)	(0.18)	(0.32)	(0.27)	(0.38)	(0.21)
Early Adolescence	0.66	0.45	0.61	0.47	0.58	0.45	0.59	0.48
	(0.15)	(0.10)	(0.06)	(0.07)	(0.08)	(0.04)	(0.09)	(0.05)
Late Adolescence	0.69	0.48	0.63	0.46	0.64	0.47	0.69	0.47
	(0.31)	(0.10)	(0.09)	(0.06)	(0.11)	(0.05)	(0.14)	(0.06)
Young Adults	0.68	0.48	0.67	0.48	0.64	0.47	0.68	0.47
	(0.27)	(0.13)	(0.24)	(0.11)	(0.18)	(0.10)	(0.25)	(0.10)

Reported in Seconds

Table A.12. Average *Total RTV*

Block	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
	M	*M*	*M*	*M*	*M*	*M*	*M*	*M*
	(SD)	*(SD)*	*(SD)*	*(SD)*	*(SD)*	*(SD)*	*(SD)*	*(SD)*
Early Childhood	0.56	0.30	0.61	0.30	0.56	0.29	0.62	0.29
	(0.40)	(0.42)	(0.43)	(0.13)	(0.50)	(0.14)	(0.46)	(0.15)
Middle Childhood	0.43	0.23	0.55	0.27	0.45	0.24	0.61	0.29
	(0.32)	(0.14)	(0.32)	(0.13)	(0.33)	(0.10)	(0.39)	(0.12)
Late Childhood	0.60	0.33	0.44	0.19	0.46	0.25	0.53	0.23
	(0.51)	(0.31)	(0.27)	(0.12)	(0.33)	(0.14)	(0.40)	(0.13)
Early Adolescence	0.43	0.15	0.26	0.14	0.17	0.10	0.28	0.15
	(0.34)	(0.13)	(0.11)	(0.09)	(0.09)	(0.05)	(0.25)	(0.07)
Late Adolescence	0.33	0.14	0.20	0.11	0.21	0.13	0.38	0.16
	(0.42)	(0.10)	(0.13)	(0.05)	(0.14)	(0.08)	(0.37)	(0.09)
Young Adults	0.30	0.14	0.25	0.12	0.23	0.12	0.31	0.13
	(0.33)	(0.10)	(0.33)	(0.10)	(0.20)	(0.10)	(0.36)	(0.11)

Table A.13. Average *Total Efficiency*

Block	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
	M	*M*	*M*	*M*	*M*	*M*	*M*	*M*
	(SD)	*(SD)*	*(SD)*	*(SD)*	*(SD)*	*(SD)*	*(SD)*	*(SD)*
Early Childhood	87.30	143.00	76.60	136.00	76.30	113.00	67.80	90.00
	(58.20)	(65.00)	(51.90)	(55.10)	(56.20)	(54.90)	(42.40)	(50.40)
Middle Childhood	99.70	164.00	86.70	147.00	88.60	142.00	86.90	117.00
	(41.60)	(48.20)	(37.80)	(47.10)	(50.10)	(39.40)	(38.20)	(44.70)
Late Childhood	110.00	156.00	108.00	152.00	111.00	133.00	108.00	148.00
	(56.40)	(63.10)	(42.30)	(44.90)	(45.90)	(61.50)	(52.50)	(42.80)
Early Adolescence	153.00	221.00	161.00	201.00	152.00	194.00	160.00	196.00
	(42.00)	(43.50)	(18.10)	(34.20)	(59.00)	(75.80)	(24.70)	(18.80)
Late Adolescence	153.00	198.00	157.00	214.00	148.00	196.00	136.00	194.00
	(53.10)	(61.50)	(21.40)	(23.30)	(44.30)	(55.20)	(43.90)	(53.60)
Young Adults	161.00	216.00	160.00	215.00	158.00	214.00	150.00	211.00
	(59.20)	(50.60)	(49.40)	(48.80)	(53.80)	(60.30)	(56.40)	(65.90)

Table A.14. ANOVA Secondary Metrics

Metrics	**Factor**	***SS***	***df***	***F***	***p***
Omission	Age	11823	5	3.25	0.009
	Gender	492	1	0.68	0.413
	Age x Gender	1470	5	0.40	0.845
	Residuals	77901	107		
PCT	Age	0.6726	5	5.41	<.001
	Gender	0.0145	1	0.59	0.446
	Age x Gender	0.0401	5	.032	0.898
	Residuals	2.6588	107		
RT	Age	1.80332	5	7.83	<.001
	Gender	0.00465	1	0.10	0.751
	Age x Gender	0.13452	5	0.58	0.712
	Residuals	1.92840	107		
RTV	Age	1.6412	5	4.74	<.001
	Gender	0.0687	1	0.99	0.321
	Age x Gender	0.1252	5	0.03	0.874
	Residuals	7.4078	107		
Efficiency	Age	120759.00	5	10.47	<.001
	Gender	16.60	1	0.01	0.933
	Age x Gender	6343.70	5	0.55	0.738
	Residuals	246807.40	107		

Table A.15. Pairwise Comparisons- *Total Omission Score*

		Middle Childhood	Late Childhood	Early Adolescence	Late Adolescence	Early Adulthood
Early Childhood	Mean Difference	-16.67	-20.61	-32.67	-29.02	-31.47
	t	-1.45	-2.04	-2.78	-2.94	-3.71
	p	0.696	0.326	0.069	0.045	0.004
Middle Childhood	Mean Difference		-3.94	-16.00	-12.35	-14.80
	t		-0.36	-1.27	-1.14	-1.55
	p		0.999	0.799	0.863	0.634
Late Childhood	Mean Difference			-12.06	-8.41	-10.86
	t			-1.07	-0.90	-1.39
	p			0.893	0.945	0.732
Early Adolescence	Mean Difference				-3.65	-1.20
	t				-0.33	-0.12
	p				0.999	1.000
Late Adolescence	Mean Difference					-2.45
	t					-0.33
	p					1.000

Table A.16. Pairwise Comparisons- *Total PCT*

		Middle Childhood	Late Childhood	Early Adolescence	Late Adolescence	Early Adulthoo d
Early Childhood	Mean Difference	-12.00	-15.00	-22.00	-22.00	-23.00
	t	-1.86	-2.70	-3.34	-3.89	-4.84
	p	0.430	0.084	0.014	0.002	<.001
Middle Childhood	Mean Difference		-3.00	-7.06	-9.94	-11.45
	t		-0.53	-1.07	-1.57	-2.05
	p		0.995	0.89	0.618	0.321
Late Childhood	Mean Difference			-7.06	-6.56	-8.06
	t			-1.07	-1.21	-1.77
	p			0.89	0.833	0.491
Early Adolescence	Mean Difference				-0.5	-1.00
	t				-0.08	-0.17
	p				1.000	1.000
Late Adolescence	Mean Difference					-1.51
	t					-0.34
	p					0.999

Table A.17. Pairwise Comparisons- *Total RT*

		Middle Childhood	Late Childhood	Early Adolescence	Late Adolescence	Early Adulthood
Early Childhood	Mean	0.05	0.11	0.36	0.28	0.32
	Difference	0.61	1.38	3.90	3.64	4.75
	t	0.990	0.742	0.002	0.005	<.001
	p					
Middle Childhood	Mean		0.05	0.30	0.23	0.26
	Difference		0.62	3.10	2.68	3.48
	t		0.989	0.029	0.089	0.009
	p					
Late Childhood	Mean			0.25	0.17	0.2097
	Difference			2.84	2.37	3.38
	t			0.059	0.175	0.013
	p					
Early Adolescence	Mean				0.07	0.045
	Difference				0.90	0.575
	t				0.946	0.992
	p					
Late Adolescence	Mean		-			0.034
	Difference					0.57
	t					0.993
	p					

Table A.18. Pairwise Comparisons- *Total RTV*

		Middle Childhood	Late Childhood	Early Adolescence	Late Adolescence	Early Adulthood
Early Childhood	Mean Difference	0.06	0.11	0.31	0.240	0.31
	t	0.56	1.13	2.79	2.56	3.76
	p	0.993	0.870	0.067	0.114	0.003
Middle Childhood	Mean Difference		0.05	0.25	0.17	0.24
	t		0.45	2.16	1.75	2.71
	p		0.998	0.287	0.498	0.081
Late Childhood	Mean Difference			0.205	0.13	0.06
	t			1.89	1.48	0.934
	p			0.410	0.675	0.937
Early Adolescence	Mean Difference				-0.07	-0.01
	t				-0.70	-0.09
	p				0.981	1.000
Late Adolescence	Mean Difference					0.065
	t					0.93
	p					0.937

Table A.19 Pairwise Comparisons- Age Group *Total Efficiency*

		Middle Childhood	Late Childhood	Early Adolescence	Late Adolescence	Early Adulthood
Early Childhood	Mean Difference	-15.805	-31.260	-86.128	-76.304	-85.890
	t	-0.77	-1.74	-4.15	-4.34	-5.69
	p	0.972	0.509	0.001	<.001	<.001
Middle Childhood	Mean Difference		-15.455	-70.324	-60.500	-70.086
	t		-0.79	-3.15	-3.14	-4.12
	p		0.969	0.025	0.026	0.001
Late Childhood	Mean Difference			-54.869	-45.045	-54.631
	t			-2.73	-2.72	-3.93
	p			0.078	0.080	0.002
Early Adolescence	Mean Difference				-9.824	-0.238
	t				0.50	0.01
	p				0.996	1.000
Late Adolescence	Mean Difference					-9.586
	t					-0.72
	p					0.980

Table A.20 Correlations Across GNG Screen Metrics

		Commission Total	Omission Total	Total RT	Total RTV
Commission Total	Pearson's r	—			
	p-value	—			
Omission Total	Pearson's r	0.023	—		
	p-value	0.803	—		
Total RT	Pearson's r	0.042	-0.779	—	
	p-value	0.654	< .001	—	
Total RTV	Pearson's r	-0.054	-0.779	0.876	—
	p-value	0.557	< .001	< .001	—

Table A.21 Dimensional Analyses—GNG Screen

% Variance Explained by Measure	% Variance Explained by 1st Contrast	Eigen Value of first Contrast
37.6%	3.0%	2.6

Table A.22 Dimensional Analyses By Block

	% Variance Explained by Measure	% Variance Explained by 1st Contrast	Eigen Value of 1st Contrast
Block 1	23.4%	15.7%	1.8
Block 2	21.4%	14.4%	1.6
Block 3	20.9%	14.2%	1.6
Block 4	27.2%	13.9%	1.7
Block 5	32.9%	11.9%	1.6
Block 6	38.3%	11.7%	1.7
Block 7	29.7%	15.0%	1.9
Block 8	30.6%	14.0%	1.8

Table A.23 EFA Factor Loadings

	Factor	
	1	**Uniqueness**
Block 1	0.788	0.380
Block 2	0.762	0.419
Block 3	0.826	0.318
Block 4	0.900	0.190
Block 5	0.802	0.357
Block 6	0.907	0.177
Block 7	0.887	0.212
Block 8	0.860	0.261

Table A.24. Model Fit Measures

	RMSEA 90% CI				Model Test		
RMSEA	**Lower**	**Upper**	**TLI**	**BIC**	**χ^2**	**df**	**p**
0.182	0.148	0.220	0.875	3.65	99.2	20	< .001

Table A.25 EFA Fixed 2-Factor Loadings

	Factor	
	1	2
Block 1	0.473	
Block 2	0.904	
Block 3	0.536	0.388
Block 4	0.805	
Block 5	0.991	
Block 6	0.387	0.386
Block 7	0.301	0.571
Block 8	1.030	

Table A.26 EFA Fixed 3-Factor Loadings

	Factor		
	1	2	3
Block 1	0.936		
Block 2	0.616		
Block 3	0.441		0.383
Block 4	0.961		
Block 5	0.972		
Block 6	0.301		0.362
Block 7	0.383	0.581	
Block 8	1.034		

Table A.27 Average Item Difficulty in Logits

	Mean Difficulty (SD)
Block 1	-3.72 (0.65)
Block 2	-3.36 (0.74)
Block 3	-3.11 (0.71)
Block 4	-2.82 (0.93)
Block 5	-2.75 (0.83)
Block 6	-2.27 (0.66)
Block 7	-2.36 (0.56)
Block 8	-2.14 (0.30)

Table A.28 One-Way ANOVA—Average Block Difficulty

	F	*df1*	*df2*	*p*
B	5.87	7	64	< .001

Table A.29—Post Hoc Analyses Average Block Difficulty

	Block	1	2	3	4	5	6	7	8
1	Mean difference	—	-0.361	-0.615	-0.901	-0.9738	-1.497 ***	-1.356 **	-1.5791 ***
	p-value	—	0.955	0.573	0.128	0.076	< .001	0.003	< .001
2	Mean difference		—	-0.254	-0.540	-0.6129	-1.136 *	-0.995	-1.2182 **
	p-value		—	0.994	0.721	0.578	0.020	0.065	0.010
3	Mean difference			—	-0.286	-0.3584	-0.882	-0.741	-0.9638
	p-value			—	0.988	0.956	0.146	0.333	0.082
4	Mean difference				—	-0.0728	-0.596	-0.455	-0.6781
	p-value				—	1.000	0.612	0.860	0.447
5	Mean difference					—	-0.523	-0.382	-0.6053
	p-value					—	0.752	0.939	0.593
6	Mean difference						—	0.141	-0.0821
	p-value						—	1.000	1.000
7	Mean difference							—	-0.2231
	p-value							—	0.997
8	Mean difference								—
	p-value								—

Note. * p < .05, ** p < .01, *** p < .001

Table A.30 Average Infit and Outfit Across Blocks

	Mean Infit (SD)	Mean Outfit (SD)
Block 1	1.15	0.90
	(0.17)	(0.44)
Block 2	1.14	1.45
	(0.12)	(1.47)
Block 3	1.05	1.55*
	(0.16)	(1.21)
Block 4	0.99	0.91
	(0.18)	(0.55)
Block 5	1.01	1.45
	0.12)	(0.88)
Block 6	0.84	0.74
	(0.17)	(0.29)
Block 7	0.98	0.93
	(0.21)	(0.44)
Block 8	0.88	0.81
	(0.10)	(0.32)

**Falls Outside of Recommended Range*

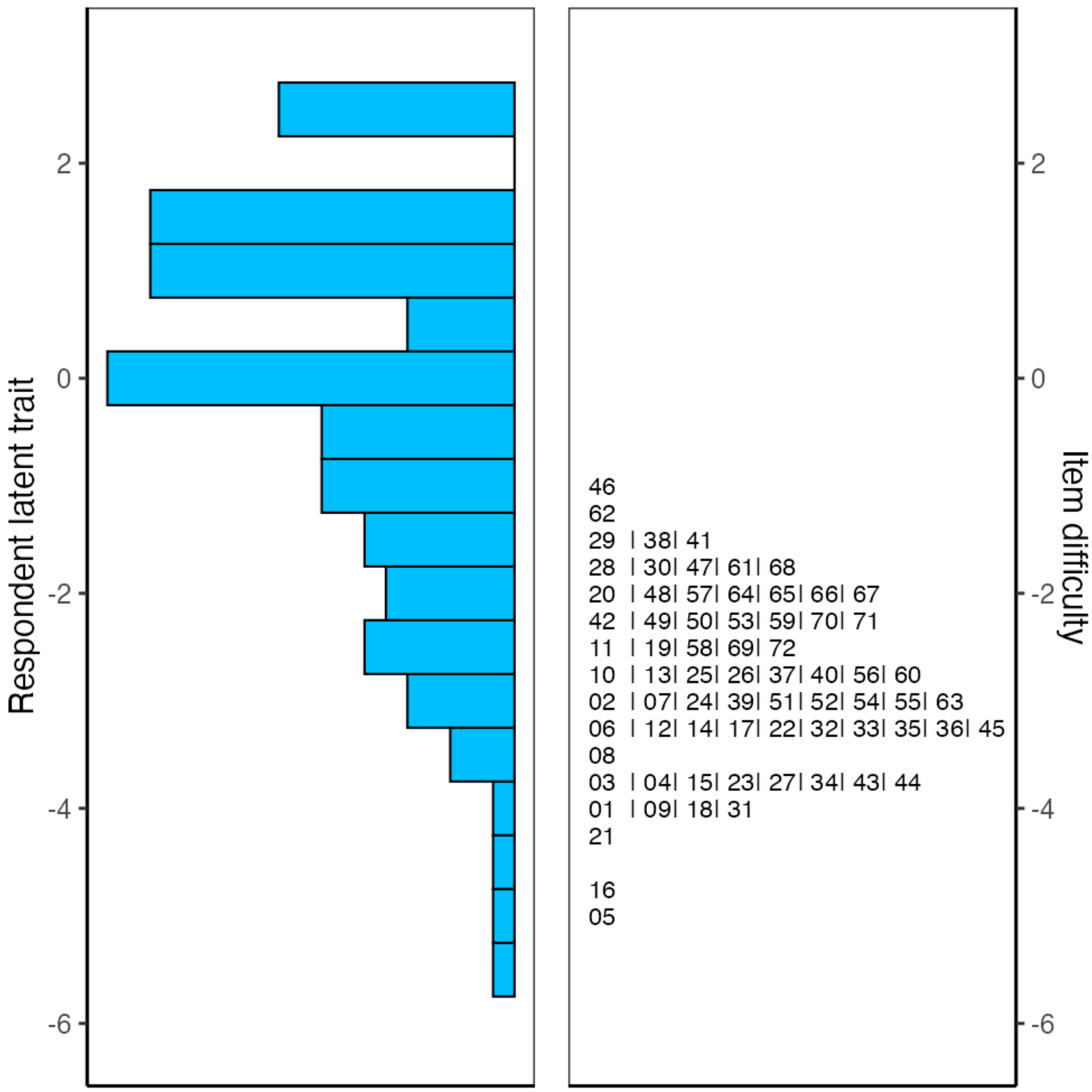


Figure A.1. Wright Map
Note: No-Go Items Labeled By Presentation Order: Block 1: 1-9; Block 2: 10-18; Block 3: 19-27; Block 4: 28-36; Block 5:37-45; Block 6: 46-54; Block 7: 55-64; Block 8: 63-72