The Pennsylvania State University

The Graduate School

AHEAD OF THE (ROC) CURVE: A STATISTICAL APPROACH TO UTILIZING EX-GAUSSIAN PARAMETERS OF REACTION TIME IN DIAGNOSING ADHD ACROSS THREE DEVELOPMENTAL PERIODS

A Dissertation in

Psychology

by

Hilary Galloway-Long

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 2020

The dissertation of Hilary Galloway-Long was reviewed and approved by the following:

Cynthia Huang-Pollock
Professor of Psychology
Dissertation Adviser
Chair of Committee

Karen Bierman
Professor of Psychology

Peter Arnett
Professor of Psychology

Barbara Schaefer
Associate Professor of Education

Kristin Buss
Professor of Psychology
Department Head, Penn State Department of Psychology

ABSTRACT

Extensive research has examined possible etiological roots of ADHD, with particular attention paid to deficits in processing speed and executive functions. However, because only about 30% to 50% of children with ADHD show deficits on neuropsychological tests relative to normative samples, one suggestion has been that current methods of measuring cognition may fail to accurately capture and describe the deficits seen within this population. The current study aimed to determine if ex-Gaussian parameters of reaction time are more accurately able discriminate between individuals with and without ADHD compared to traditional methods.

Cognitive task performance was evaluated in individuals with and without ADHD across three developmental periods: kindergarten-age, middle childhood, and early adulthood. Receiver Operating Characteristic (ROC) curves and the area under the curve (AUC) were used to examine the ability of Go/No-Go and SSRT task parameters to accurately identify individuals meeting diagnostic criteria for ADHD. Possible clinical cut-offs were also explored. Findings were consistent across samples such that the traditionally used variables (SSRT and failed inhibit rate), as well as variable reaction time (indexed by SDRT) successfully discriminated between individuals with and without ADHD. When using ex-Gaussian parameters, this was generally found to be driven by lapses in performance (indexed by tau). Though no single variable was found to consistently provide the greatest combination of sensitivity and specificity when using Youden's Index, optimal cutoffs generally showed higher specificity than sensitivity.

Findings suggest that reaction time can identify individuals with ADHD as well as traditional measures such as SSRT and failed inhibit rate, and that using ex-Gaussian parameters provides a better description of variability in performance during speeded tasks than mean/SD. Therefore, assessment tools would benefit from implementing these methods.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

This dissertation was made possible through multiple funding sources. The FRIENDS Study was funded by National Institute of Mental Health grant 1R34MH085889 to Karen Bierman and Cynthia Huang-Pollock; the CAL Study was funded by National Institute of Mental Health Grant R01 MH084947 to Cynthia Huang-Pollock; and the Brain and Behavior Study was funded by National Institutes of Health grants UL1 TR002014 and KL2 TR002015 and the Brain and Behavior Foundation grant 25004. The content of this dissertation is solely the responsibility of the author and does not necessarily represent the official views of the National Institute of Mental Health, the National Institutes of Health, or the Brain and Behavior Foundation.

Introduction

Attention Deficit Hyperactivity Disorder (ADHD) is a complex and heterogeneous disorder, characterized by difficulties with sustaining attention, hyperactivity, and impulsivity (American Psychiatric Association, 2013). Affecting an estimated 3-7% of children in the United States, ADHD is the most common of childhood psychiatric diagnoses, and as such, represents one of the largest sources of mental health service costs and academic underachievement (Pelham, Foster, & Robb, 2007). Beginning in early childhood and identified by parents' and teachers' behavioral observations as children begin school, ADHD is associated with numerous negative outcomes including poor academic achievement, and higher likelihood of repeating a grade in school (Biederman et al., 2004a; Fried et al., 2016). But the challenges associated with ADHD affect life outside of school as well. Children and adolescents with ADHD have poorer social skills, higher rates of risky or dangerous behaviors, and more stressful parental relationships than children without ADHD (DuPaul, McGoey, Eckert, & VanBrakle, 2001; Gudjonsson, Sigurdsson, Sigfusdottir, & Young, 2012). Given these negative outcomes, it is imperative that the roots of the disorder be better understood so that assessment and intervention might improve.

Therefore, an essential avenue within current research has been to explore specific factors which may function as causal mechanisms in the development of ADHD. Possible contributors to ADHD have included neurologic abnormalities (Fallgatter, Ehlis, & Herrmann, 2004), genetic risk (Biederman et al., 1992; Gallo & Posner, 2016; Nigg, Nikolas, & Burt, 2010), tobacco and lead exposure (Braun, Kahn, Froehlich, Auinger, & Lanphear, 2006; Froehlich et al., 2009; Nigg, Nikolas, Knottnerus, Cavanagh, & Friderici, 2010), reward dysfunction (Solanto, Wender, & Bartell, 1997; Volkow et al., 2011), shortened delay discounting gradients (Scheres, Tontsch,

Thoeny, & Kaczkurkin, 2010; Wilson, Mitchell, Musser, Schmitt, & Nigg, 2011), and impaired executive functioning (EF); (Barkley, 1997; Nigg, Willcutt, Doyle, & Sonuga-Barke, 2005). Of these, executive dysfunction has arguably received the most research attention (e.g., Bellgrove, Hawi, Gill, & Robertson, 2006; Jonas & Markon, 2014; Kuntsi, Oosterlaan, & Stevenson, 2001; Pauli-Pott, Dalir, Mingebach, Roller, & Becker, 2013)

Despite theoretical support for the theory of EF as a potential causal mechanism of ADHD, group effect sizes are small to moderate (see Boonstra, Oosterlaan, Sergeant, & Buitelaar, 2005; Huang-Pollock, Karalunas, Tam, & Moore, 2012; Huang-Pollock & Nigg, 2003; Lijffijt, Kenemans, Verbaten, & van Engeland, 2005; Oosterlaan, Logan, & Sergeant, 1998; Pauli-Pott & Becker, 2011; Schwartz & Verhaeghen, 2008; Willcutt, Doyle, Nigg, Faraone, & Pennington, 2005). These small to moderate effect sizes have been interpreted as suggestive of population-level heterogeneity, in which some (but by no means all) children with ADHD demonstrate an EF deficit (Karalunas et al., 2017; Wahlstedt, Thorell, & Bohlin, 2009). In fact, only a minority of children (roughly 30%-50%) diagnosed with ADHD show deficits on at least one task of executive function (Nigg et al., 2005). However, it may *also* be that current methods of measuring cognition fails to capture the deficit in all who possess it.

The current study focuses specifically on processing speed, as indexed by reaction time (RT) to simple and forced choice RT tasks. Processing speed is understood to function as a base-level mechanism which influences higher-level functions, most notably working memory (see: Barrouillet, Gavens, Vergauwe, Gaillard, & Camos, 2009; Bayliss, Jarrold, Baddeley, Gunn, & Leigh, 2005; Salthouse, 1991). Extensive evidence associates impairments in processing speed with ADHD (see: Castellanos & Tannock, 2002; Epstein et al., 2011b; Fry & Hale, 2000; Galloway-Long & Huang-Pollock, 2018; Hervey, Epstein, Curry, Tonev, Eugene Arnold, Keith

Conners, et al., 2006; Jacobson et al., 2011; Kail & Salthouse, 1994; Kalff et al., 2005; Karalunas & Huang-Pollock, 2013; Leth-Steensen, Elbaz, & Douglas, 2000; Magimairaj & Montgomery, 2012; Martinussen, Hayden, Hogg-Johnson, & and Tannock, 2005; Salthouse, 1996).

That being said, the ways in which performance on speeded reaction time tasks is described can and should be more precise. In particular, reaction times are not normally distributed, despite the fact that mean and standard deviation remain the most commonly used method to describe speeded performance. Instead, reaction time distributions are more accurately described with the ex-Gaussian distribution. An ex-Gaussian distribution is formed by integrating normal and exponential distributions, to accurately reflect distributions that do not have a lower tail, but do have a long upper tail (Dawson, 1988; Ratcliff & Murdock, 1976). It is described by the parameters mu (which is the mean of the normal portion of the distribution), sigma (which is the standard deviation of the normal portion of the distribution; representing variability or spread), and tau (which is the mean of the exponential tail of the distribution; representing positive skew due to relatively few exceedingly slow trials) (Dawson, 1988; Lacouture & Cousineau, 2008; Van Zandt & Townsend, 2014). See Figure 1 for illustration. Substantial literature supports claims that the commonly found slower/more variable performance on speeded reaction time tasks among children with ADHD is not due to slower mean/mu or larger standard deviation/sigma, but that it is driven primarily by increased tau (Buzy, Medoff, & Schweitzer, 2009; Epstein et al., 2011a; Gu, Gau, Tzang, & Hsu, 2013; Karalunas & Huang-Pollock, 2013; Leth-Steensen et al., 2000; Shahar, Teodorescu, Usher, Pereg, & Meiran, 2014). Further, replicated evidence indicates that larger tau specifically contributes to reduced working memory capacity and reduced inhibitory control (Buzy et al., 2009; Epstein et al., 2011a; Gu et al., 2013; Karalunas & Huang-Pollock, 2013; Leth-Steensen et

al., 2000; Schmiedek, Oberauer, Wilhelm, Suss, & Wittmann, 2007; Shahar et al., 2014).

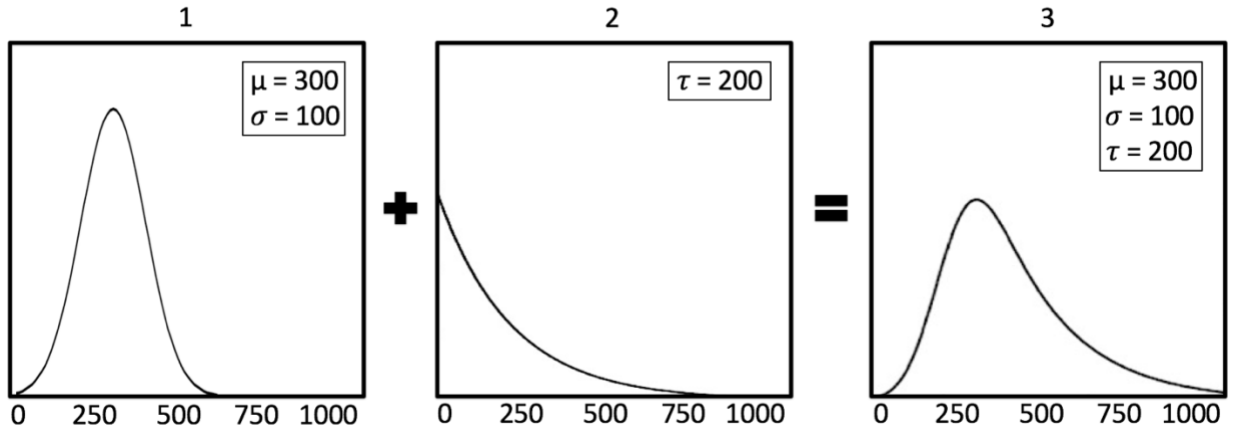

Figure 1. Functions for 1) a Gaussian (normal) distribution with mu = 300 ms, sigma = 100 ms; 2) an exponential distribution with tau = 200 ms, and the resulting 3) combined ex-Gaussian distribution with mu=300ms, sigma= 100 ms, and tau = 200 ms.

One of the driving forces in identifying potential cognitive mechanisms associated with ADHD is improving assessment. More precisely measuring cognition and establishing more accurate group-level effect sizes is critical to this endeavor. However, the ultimate translational goal of this type of research is to improve the field's ability to predict and detect the presence of disorder. The current study aims to determine if ex-Gaussian parameters of reaction time are more accurately able discriminate between children and young adults with and without ADHD compared to traditional methods of indexing cognition that assume a normal distribution of performance. In doing so, the study may contribute to the field's current understanding of causal mechanisms associated with the disorder, and highlight ways in which neuropsychological evaluations may be improved by implementing more precise assessment methods. Before discussing the specific hypotheses and methods of this study, I will discuss the potential of executive functions serving as an endophenotype of the disorder, the role processing speed plays in executive functions and in ADHD, and the methods available to measure processing speed.

Finally, I will discuss the primary method used to evaluate the diagnostic discriminability of assessment measures and the findings of the few studies which have evaluated the ability of common cognitive assessments to identify individuals with ADHD.

1. Executive Functions as an endophenotype of ADHD

At the group level, a wide range of cognitive and EF deficits have been commonly observed in children with ADHD (Barkley, 1997; Nigg & Casey, 2005). Significant group differences have been found for sustained attention/vigilance (Bellgrove et al., 2006; Huang-Pollock et al., 2012), inhibition (Crosbie, Perusse, Barr, & Schachar, 2008; Crosbie & Schachar, 2001; Schachar et al., 2004), delay aversion (Bitsakou, Psychogiou, Thompson, & Sonuga-Barke, 2009; Kuntsi, Rogers, et al., 2006), and processing speed and working memory (Epstein et al., 2011b; Fry & Hale, 2000; Hervey, Epstein, Curry, Tonev, Eugene Arnold, Keith Conners, et al., 2006; Kail & Salthouse, 1994; Kalff et al., 2005; Karalunas & Huang-Pollock, 2013; Karalunas, Huang-Pollock, & Nigg, 2012; Leth-Steensen et al., 2000; Magimairaj & Montgomery, 2012; Nigg et al., 2018; Salthouse, 1996). Given the relative consistency with which EF deficits have been identified in children with ADHD, some have suggested that executive dysfunction may serve as a possible endophenotype, contributing to our understanding of the etiology of the broader phenotype or behavioral profile associated with ADHD (Bellgrove et al., 2006; Crosbie et al., 2008; Doyle, Biederman, Seidman, Weber, & Faraone, 2000; Doyle et al., 2005; Pauli-Pott & Becker, 2011).

A phenotype, defined as the collection of observable behavioral characteristics of an organism, is recognized to be resulting from both genetic and environmental influences, and in psychology refers to the highly complex behavioral or symptom profiles composing a given clinical syndrome or diagnosis (Gottesman & Gould, 2003). While the classification of

psychological phenotypes is valuable in grouping together broad swathes of individuals presenting with generally similar profiles, their overall complexity and the likely involvement of multiple biological, genetic, and environmental influences suggests that each phenotype may be reached through multiple etiological pathways (Gottesman & Gould, 2003; Kebir & Joober, 2011). As such, research attempting to elucidate the casual mechanisms of a psychiatric disorder is unlikely to be successful if relying exclusively on phenotypic categorization.

In contrast, an endophenotype is defined as an intermediate-level phenotype or collection of traits which is generally simpler or more specific (Gottesman & Gould, 2003; Kebir, Tabbane, Sengupta, & Joober, 2009). To qualify as an endophenotype, the construct must be a) significantly associated with the disorder, b) heritable, c) present even if symptoms associated with the disorder are in remission, d) more prevalent in family members with than without the disorder, e) more prevalent in family members without the disorder than the general population, and f) can be reliably measured (Gottesman & Gould, 2003; Lenzenweger, 2013). Given such strict criteria, successfully identifying an endophenotype of a psychiatric disorder therefore holds promise for identifying a causal mechanism influencing a subset of individuals with the disorder. Within research on causal mechanisms of ADHD, significant attention has been paid to the possibility of deficits in executive functions as an endophenotype.

Numerous cognitive domains have been suggested as possible endophenotypes of ADHD, the most widely studied including: sustained attention, impulsivity, delay aversion, inhibition, and working memory (Bellgrove et al., 2006; Boonstra, Kooij, Oosterlaan, Sergeant, & Buitelaar, 2010; Crosbie et al., 2013; Johnson et al., 2008; Jonas & Markon, 2014; Kieling, Roman, Doyle, Hutz, & Rohde, 2006; Kuntsi & Stevenson, 2001; Nigg et al., 2018; Pauli-Pott et al., 2013). However, group effect sizes remain modest (Benca et al., 2017; Nigg et al., 2005;

Oosterlaan et al., 1998; Willcutt, Pennington, Olson, Chhabildas, & Hulslander, 2005), leading many in recent years to argue that the potential of EF to serve as a cognitive endophenotype to the disorder has not lived up to its promise. More recently, a familial study examining working memory and inhibition found that while significant familial associations were present during childhood, they were no longer evidenced by adolescence, and suggested that while both ADHD and EF seem to share influences during childhood, they are largely independent by adolescence (Thissen et al., 2014). Two possible explanations for this continued lack of clarity have been proposed.

First, many researchers have proposed that because both the behavioral expression of ADHD and executive functions are such broad constructs, it's likely that there exists a great deal of within-group heterogeneity, meaning that while some individuals will show deficits in one subdomain but perform within normal ranges on others, other individuals may show an opposite range of skills and deficits. For example, Wahlstedt et al. (2009) found differing neurocognitive groups based on symptom presentation. Similarly, other data-driven analytical techniques including community detection and latent class analysis of EF task performance have found that children with ADHD could be grouped according to differing cognitive profiles (Karalunas et al., 2017), which may mirror similar profiles in a typically developing population (Fair, Bathula, Nikolas, & Nigg, 2012). However, this line of inquiry is relatively recent, and replication attempts in independent samples have not yet been conducted to provide consensus on the identity of these subgroups or subtypes.

Second, despite the likely possibility of genuine heterogeneity in the types of EF deficits associated with ADHD, evidence also suggests that significant flaws exist in the standard methods used to measure EF (Cepeda, Blackwell, & Munakata, 2013; Doyle et al., 2000;

Edwards et al., 2007; Golden et al., 1981; Homack & Riccio, 2004; Karalunas, Bierman, & Huang-Pollock, 2016; Kuntsi, Neale, Chen, Faraone, & Asherson, 2006; Nigg et al., 2005; Reynolds & Horton, 2014; Snow & Hynd, 1984; Willcutt, Doyle, et al., 2005). These include poor or otherwise unknown test-retest reliability (Karalunas et al., 2016; Kuntsi, Neale, et al., 2006; Willcutt, Doyle, et al., 2005), imprecise tasks which involve multiple neurocognitive processes (Golden et al., 1981; Willcutt, Doyle, et al., 2005), and insufficient sensitivity or specificity to properly identify deficits (Cepeda et al., 2013; Doyle et al., 2000; Edwards et al., 2007; Homack & Riccio, 2004; Nigg et al., 2005). Though generally adequate reliability has been established for many cognitive measures in middle-childhood through adulthood (Willcutt, Doyle, et al., 2005), values are moderate and relatively less is known regarding the reliability of EF tasks during early childhood (Karalunas et al., 2016). Additionally, because many traditional neuropsychological tests of EF were developed to identify effects of traumatic brain injury (Golden et al., 1981), they are most valid and reliable when identifying broad deficits. However, they deliberately require the simultaneous recruitment of multiple cognitive processes, which limits their ability to parse apart those specific functions when examining more nuanced cognitive impairments such as those associated with neurodevelopmental disorders (Reynolds & Horton, 2014; Snow & Hynd, 1984; Willcutt, Doyle, et al., 2005). Finally, research indicates that many common neuropsychological assessments have relatively low sensitivity and specificity, and are often unable to accurately discriminate between children with and without ADHD (Cepeda et al., 2013; Doyle et al., 2000; Edwards et al., 2007; Homack & Riccio, 2004; Nigg et al., 2005). Together, these concerns suggest that continuing to focus on EF broadly and relying on limited methodologies may be masking consistent cognitive dysfunction throughout the ADHD population.

One way forward may be to focus on a cognitive marker that is both associated with executive functioning and also more tractable and easily defined, thereby facilitating the mechanistic link between genetic contributions to the disorder and the disorder itself. A recent review found that high reaction time variability was the cognitive marker most consistently associated with the candidate genes DRD4 and DAT1 (Kebir et al., 2009), and a recent study of the association between cognitive functions and candidate genes found a strong association between the HTR2A gene and reaction time variability (Pinto, Asherson, Ilott, Cheung, & Kuntsi, 2016). Given this evidence of biological markers for impairments in reaction times tasks, the current study will further discuss the role of variability in processing speed as a more precise marker of cognitive dysfunction in ADHD, how methods of measurement can and should be improved to increase accuracy, and test the sensitivity and specificity of these improved measurement methods across multiple age groups.

2. Processing Speed

Processing speed is a critical construct because it is thought to function as a base-level mechanism upon which other higher level functions (e.g., working memory) rely (Barrouillet et al., 2009; Bayliss et al., 2005; Ferguson & Bowey, 2005; Gaillard, Barrouillet, Jarrold, & Camos, 2011; Kail, 1992; Kail & Park, 1994; Kail & Salthouse, 1994; Salthouse, 1991, 1996; Vicari, Caravale, Carlesimo, Casadei, & Allemand, 2004), and because impaired processing speed is often associated with neurocognitive disorders such as ADHD (Adamo et al., 2014; Borella, de Ribaupierre, Cornoldi, & Chicherio, 2013; Buzy et al., 2009; Castellanos & Tannock, 2002; Epstein et al., 2011a; Galloway-Long & Huang-Pollock, 2018; Hervey, Epstein, Curry, Tonev, Eugene Arnold, Conners, et al., 2006; Jacobson et al., 2011; Kalff et al., 2005; Karalunas et al., 2012; Kofler et al., 2013; Kuntsi et al., 2001; Leth-Steensen et al., 2000; Martinussen et al.,

2005; Tye et al., 2016; van Belle, van Hulst, & Durston, 2015; Weigard & Huang-Pollock, 2016).

Common methods of measuring PS. Extensive research within both typically developing and clinical populations have used paper-pencil tasks of processing speed to identify mechanisms explaining deficits in higher-order constructs including working memory. Some of the earliest studies within normative populations found that faster processing speed, indexed using the WISC subscales, significantly predicted faster and more accurate recall in complex verbal and non-verbal working memory span tasks (Kail, 1992; Kail & Park, 1994; Kail & Salthouse, 1994; Salthouse, 1991, 1996). Similar results have also been found in studies of children with ADHD (Jacobson et al., 2011; Mealer, Morgan, & Luscomb, 1996). This association was thought to suggest that more efficient processing improved working memory by providing more time to employ techniques such as rapid rehearsal to maintain and refresh remembered items (Fry & Hale, 2000; Kail, 1992; Kail & Park, 1994).

Though these tasks are simple and fast to administer, and can quickly provide a general estimate of an individual's speed and accuracy, their methodology also results in several limitations. First, in addition to information processing speed, these tasks also require strong fine motor skills, visual scanning, short term memory, and effective use of strategies. Second, because speed is measured by an administrator using a stopwatch while observing the individual's performance, it is subject to a level of imprecision. Third, because most tasks include only one total completion time for the entire task, no observations can be made about the consistency or variability of speed throughout the task. More precise and detailed methods are therefore necessary.

Computerized Tasks. Studies operationalizing processing speed using computerized

reaction time tasks with normative populations found that faster RTs predict better performance on tasks of verbal (Barrouillet et al., 2009; Bayliss et al., 2005; Ferguson & Bowey, 2005; Gaillard et al., 2011) and visuo-spatial (Bayliss et al., 2005; Vicari et al., 2004) working memory. In addition to predicting WM capacity in typically developing individuals, mean reaction time (MRT) has also been found to be associated with deficits in WM and other EF domains in children with ADHD. Additionally, children with ADHD have consistently shown slower MRT (Alderson, Rapport, & Kofler, 2007; Castellanos et al., 2005; Lijffijt et al., 2005; Oosterlaan et al., 1998), and the association remains significant when the influence of age is controlled (Nigg, 1999). However, studies in which effect sizes for the standard deviation of reaction time (SDRT) were larger than those for MRT (Alderson et al., 2007; Lijffijt et al., 2005) suggest that individual variability in speed may be a more central core feature of ADHD than behavioral inhibition or other EFs.

These outcome variables are certainly informative, and can more greatly contribute to the broader picture of a neuropsychological assessment than paper-and-pencil tasks. However, analyses of these tasks could still be improved. This is because while studies of RT traditionally rely on mean and standard deviation, this practice relies on the false assumption that RT distributions are normal; ignoring the distributions' positive skew, and consequently, the theoretical interpretations of slower or inconsistent speed.

Intraindividual Variability in Processing Speed. Many common constructs used within psychological research, including standardized scaled scores from behavioral questionnaires and neuropsychological measures, are represented using a normal (Gaussian) distribution in which the mean is considered the best index of central tendency, and skew towards larger or smaller values is thought to be minimal (Dawson, 1988; Van Zandt &

Townsend, 2014). This convention is also frequently applied to many tasks of reaction time, which predominantly rely on mean and standard deviation to describe overall speed of performance (e.g., Conners, 2004). However, reaction time data is not normally distributed, as it is bound by a fastest possible response time of zero milliseconds, but potentially infinitely slow responses (Dawson, 1988; Van Zandt & Townsend, 2014). As such, RT is best represented by the non-normal ex-Gaussian distribution, which integrates an exponential distribution with a normal distribution in a way that accounts for the positive skew caused by substantially more exceptionally slow RTs than fast. The parameters mu and sigma refer to the mean and standard deviation of the normal portion of the distribution; tau characterizes the mean of the exponentially shaped (skewed) portion of the distribution (Lacouture & Cousineau, 2008). See Figure 1 for illustration.

Beyond simply describing the shape of the reaction time distribution more accurately than mean and standard deviation, further examination of ex-Gaussian parameters may contribute to a better understanding of what aspects of processing speed are associated with other cognitive domains, thereby helping to explain why children with ADHD perform more poorly on speeded RT tasks and tasks of EF. This is because information processing, as operationalized within commonly used speeded response tasks (such as CPT, GNG, SSRT tasks), is understood to entail multiple broad sub-processes. These processes include perceptual encoding, decision-making, and fine-motor output (Antonini, Narad, Langberg, & Epstein, 2013; Luce, 1986; Ratcliff & Smith, 2010; Salthouse, 1996; A. Voss, Nagler, & Lerche, 2013; White, Ratcliff, & Starns, 2011). Some researchers have argued that these components may be inferred from the shape of the RT distribution that is produced during speeded task performance (Luce, 1986; Myerson, Hale, Zheng, Jenkins, & Widaman, 2003; Rotello & Zeng, 2008; Smith & Ratcliff,

2004). Therefore, individual differences in overall speed (mu), general variability (sigma), or the size of a distribution's skew (tau) are important because they may capture unique sources of variance in the relationships between processing speed and higher level cognitive functions and with ADHD.

Competing theories have been proposed associating each ex-Gaussian parameter with particular psychological constructs (see Matzke & Wagenmakers, 2009 for a review). One particularly well-received and common argument suggests that because motor preparation/ execution and informational encoding are relatively automatic functions, the general speed or efficiency of these processes is relatively normally distributed and is best measured by the indices of central tendency (i.e., mu and sigma) (Abney, McBride, & Petrella, 2013; Balota & Spieler, 1999; Gmehlin et al., 2014; Gordon & Carson, 1990; Hockley, 1984; Luce, 1986; Madden et al., 1999; Moret-Tatay et al., 2016; Rotello & Zeng, 2008). Additionally, the more effortful or attentional processes are thought to be best described by the exponential tail of the distribution (i.e., tau) (Abney et al., 2013; Balota & Spieler, 1999; Gmehlin et al., 2014; Gordon & Carson, 1990; Hockley, 1984; Luce, 1986; Madden et al., 1999; Moret-Tatay et al., 2016; Rotello & Zeng, 2008).

The popularity of this view is in part because of a phenomenon known as the "worst performance rule" (Larson & Alderton, 1990), which indicates that the slowest RTs, rather than the fastest or the median, are the most strongly correlated with higher order cognitive processes such as IQ and executive function. The specific conceptual interpretation of tau hypothesized for the current study suggests that tau serves as an index of the speed of information accumulation during the decisional component of a RT. This interpretation is strongly informed by a mathematical model of choice reaction time task performance known as the diffusion model;

which includes parameters that have been extensively empirically validated (Ratcliff, 2002, 2014; Andreas Voss, Rothermund, & Voss, 2004). In particular, "drift rate" is a parameter from the diffusion model that indexes the speed of information or evidence accumulation during the decision-making process, and though not a perfect 1:1 association, (Matzke & Wagenmakers, 2009) has been found to be substantively negatively correlated with tau (Karalunas & Huang-Pollock, 2013).

The findings of several studies support these views. When the more accurate ex-Gaussian distribution parameters have been used to describe RT, tau (but not mu or sigma) predicted performance in WM in typically developing children and adults as well as in a range of clinical populations (Borella et al., 2013; Karalunas & Huang-Pollock, 2013; Kofler et al., 2013; Mella, Fagot, Lecerf, & de Ribaupierre, 2015; Schmiedek et al., 2007; Shahar et al., 2014; Unsworth, Redick, Lakey, & Young, 2010). Within the ADHD literature, analyses of ex-Gaussian parameters found that much of the group difference in RT is due to longer tau rather than mu (Buzy et al., 2009; Epstein et al., 2011a; Gu et al., 2013; Karalunas & Huang-Pollock, 2013; Leth-Steensen et al., 2000; Shahar et al., 2014). A recent study simultaneously examined inspection time and ex-Gaussian parameters of inspection time to parse apart perceptual encoding, motor execution, and decisional processes, and found that children with ADHD did not differ on inspection time, mu, or sigma, but did demonstrate significantly longer tau; further indicating that individuals with ADHD are characterized by generally normal performance, but with a relatively small number of exceedingly slow responses (Galloway-Long & Huang-Pollock, 2018). When not properly acknowledged, these few exceedingly slow responses lead to a biased or inaccurately inflated mean reaction time (MRT), often leading to erroneously concluding that deficits lie in slower information processing rather than impaired evidence

accumulation during decisional processes.

3. Assessing Discriminability

Various constructs may differ based on diagnostic group membership, however, the presence of diagnostic group differences alone does not inherently indicate the construct's ability to successfully discriminate between the two diagnostic groups. For example, though there is a significantly higher rate of males than females with ADHD, using an individual's status as male to predict an ADHD diagnosis would be both highly inaccurate and foolish. Given this, measures intended to contribute to diagnostic decisions must be evaluated for their diagnostic accuracy. This is done by asking, compared to the field's current diagnostic gold standard, how accurately do scores or responses on the proposed measure correctly identify the presence or absence of a disorder within an individual (Hajian-Tilaki, 2013; Youngstrom, 2014).

Key constructs used for this purpose include sensitivity and specificity and positive and negative predictive power (Hajian-Tilaki, 2018; Kraemer, 1992; Swets, 1982; Youngstrom, 2014). Sensitivity is defined as the percentage of people *with* the disorder who were *correctly identified* by the test, while specificity is defined as the percentage of people truly *without* the disorder who were *correctly not identified* by the test; see Figure 2 for illustration (Hajian-Tilaki, 2018; C. J. Smith, 2012). For a truly perfect test with complete accuracy, both sensitivity and specificity would be 100%. For a truly useless test, sensitivity and specificity would both be 50% and a coin flip would be equally clinically valuable. Importantly, sensitivity and specificity are inversely related such that a highly sensitive test which correctly identifies nearly all individuals with a disorder will not be as specific, and will therefore incorrectly identify some individuals who do not have the disorder (Hajian-Tilaki, 2018; C. J. Smith, 2012; Youngstrom, 2014). Therefore, a balance must be struck to simultaneously maximize both sensitivity and specificity

to suit the needs of the test at hand. Effectively evaluating a test's discriminability requires a method incorporating each of these values.

		Disorder	
		Positive	Negative
Test	Positive	True Positive (a)	False Positive Type I error (b)
	Negative	False Negative Type II error (c)	True Negative (d)
		Sensitivity = a / (a+c)	Specificity = d / (b+d)

Figure 2. Evaluating sensitivity and specificity of a diagnostic test compared to the standard assessment method used to determine diagnosis. Sensitivity is the number of "true positives" (i.e., correct diagnoses), divided by the number of "true positives" plus "false negatives" (i.e., missed diagnoses or Type II errors). Specificity is the number of "true negatives" (i.e., correct non-diagnosis) divided by the number of "true negatives" plus "false positives" (i.e., incorrect diagnosis or Type I errors).

Receiver Operating Curve Analyses. Though much of pediatric psychology literature continues to rely on correlations or effect sizes such as *d* (Cohen, 1988; Youngstrom, 2014), another common and well-respected method used to quantitatively evaluate the sensitivity and specificity of a test, including determining if it performs more accurately than chance, is called Receiver Operating Curve (ROC) Analyses (Hajian-Tilaki, 2013, 2018; Hajian-Tilaki, Hanley, & Nassiri, 2011; Hanley, 1989; Hanley & McNeil, 1983; Youngstrom, 2014). Modeled after Signal Detection Theory (Hanley, 1989), ROC analyses use parametric curve fitting (Hanley, 1989) and graph sensitivity on the y-axis with values ranging from 0 to 1, and graph the false alarm rate (e.g., the inverse of specificity or, 1-specificity) on the x-axis with values also ranging

from 0-1. The shape of the curve is determined by incrementally increasing the decision criterion (or score cutoff) from the minimum to the maximum score such that the full range of possible trade-offs between sensitivity and specificity is plotted (Hanley, 1989; Youngstrom, 2014). Curves farthest to the top left of the graph are therefore both highly sensitive and specific, while those closest to the diagonal line are neither, and perform close to chance (Hajian-Tilaki, 2013, 2018; Hajian-Tilaki et al., 2011; Hanley, 1989; Hanley & McNeil, 1983; Youngstrom, 2014). See Figure 3 for detail.

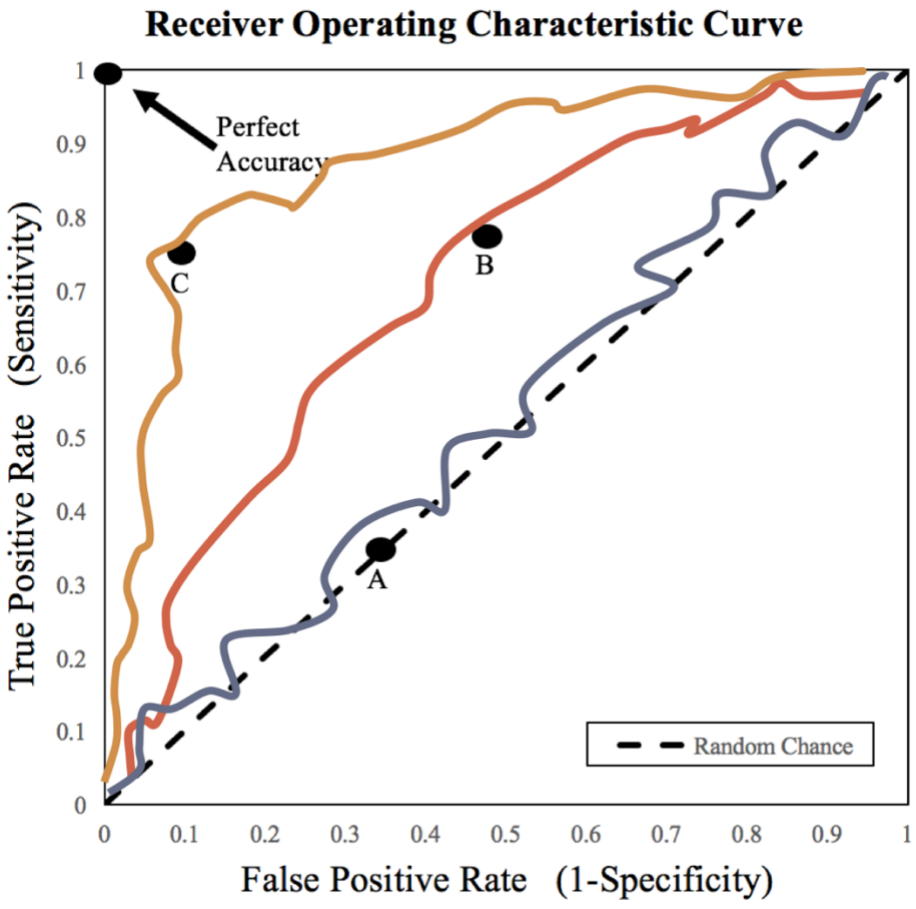

Figure 3. Diagram illustrating a Receiver Operating Curve. Y-axis represents sensitivity from 0-1, X-axis illustrates the inverse of specificity (or the false positive rate). The diagonal line represents sensitivity/specificity equal to chance. Point A represents a measure as predictive as chance and B represents a measure with moderate diagnostic value. Point C has strong diagnostic value.

Beyond the readily interpretable visual provided by the graph, ROC analyses also include the area under the curve (AUC) statistic, which quantifies each curve's distance from chance (Hanley, 1989; Youngstrom, 2014). AUC values range from 0-1, with 1 indicating perfect sensitivity and specificity, .5 accuracy equal to chance, and 0 complete failure. Statistical output included with AUC values also provides a *p* value indicating if the measure is significantly better than chance at making a correct diagnosis. Though strict benchmarks intended for use in fields such as engineering or biomedical assessment suggest that values 0.9-1.0 are ''excellent,'' .80-.89 ''good,'' 0.70-.79 ''fair,'' and <0.70 "poor", these values are likely unreasonable in the less clear-cut world of mental health; in fact, many well-supported behavioral questionnaires (e.g., the SDQ and CBCL) include AUC estimates between .70-.90 (R. Goodman & Scott, 1999; Raiker et al., 2017; Youngstrom, 2014). Multiple AUC values can also be tested against each other to determine if either serves as a stronger predictor than the other by using the AUC values, their standard errors, and the correlation between the two tests of interest (Hanley & McNeil, 1983; Youngstrom, 2014).

Selecting Cut-Offs and the Costs of Over or Under-Diagnosis. After identifying tests with significant AUCs indicating sufficient discriminability, ideal cutoff scores for clinical implementation can be determined based on the intended use of the test and the costs/benefits of diagnostic errors. Generally, diagnostic assessment includes a two-step process; one in which a low-cost/low-effort and highly sensitive screening measure identifies anyone who may possibly meet the diagnosis, followed by higher-effort but more specific measures identifying fewer individuals using more stringent criteria (Florkowski, 2008). Typically, computerized cognitive tasks such as those described previously would generally fit into this second category and

therefore would prioritize specificity. However, because the current study is largely exploratory, it will examine the sensitivities and specificities within the tasks at hand and discussion will include the contexts within which the tasks may provide the most diagnostic value.

Diagnostic errors are divided into two categories; Type I (i.e., a false positive), and Type II (i.e., a false negative) see Figure 2 for greater detail (Sheskin, 2003). Generally, the decision to minimize either type I or type II errors is determined based on which presents the greatest risk or consequence. In the case of diagnosing a child or young adult with ADHD, the possible consequences of missing a diagnosis may include the most severe outcomes associated with ADHD such as school failure or poor work productivity, dangerous or sensation-seeking behavior, illegal drug use, and increased peer and familial conflict (Biederman et al., 2004b; Birnbaum et al., 2005; Eakin et al., 2004; D. W. Goodman, 2007; Harpin, 2005; Minde et al., 2003). In contrast, the possible consequences of incorrectly providing a diagnosis may include social stigma associated with the diagnosis, possible misdiagnosis and failure to identify a more appropriate disorder, inflated healthcare costs, or even possible iatrogenic effects of an inappropriately prescribed medication (Abramovitch, 2016; Birnbaum et al., 2005; Evans, Morrill, & Parente, 2010; Wiener et al., 2012).

ROC Analyses in ADHD Diagnosis. Though a handful of studies have examined the diagnostic discriminability of computerized attention or inhibition tasks in individuals with ADHD, most have focused on outcome variables measuring accuracy or overall performance rather than processing speed or RT (Berger, Slobodin, & Cassuto, 2017; Doyle et al., 2000; Edwards et al., 2007; Faraone et al., 2016; Jiménez-Figueroa et al., 2017; Teicher, Polcari, Fourligas, Vitaliano, & Navalta, 2012). Three relatively recent studies tested the discriminability of the Conners' CPT in children (Edwards et al., 2007; Faraone et al., 2016) and adults (Teicher

et al., 2012) with somewhat lackluster results (the highest AUCs for either the overall index or failed inhibit rate were between .62 - .64). Similar studies implementing novel CPT and GNG tasks of their own design found significant discriminability, but with a wide range of AUC values (.58 - .91) and also included no specific measures of RT (Berger et al., 2017; Doyle et al., 2000; Jiménez-Figueroa et al., 2017). Only one study compared the Conners' CPT with a parent-report questionnaire and a novel GNG task called "Groundskeeper" (Faraone et al., 2016). They found that discriminability was better for both the Conner's Rating Scale (AUC = .76), and their Groundskeeper GNG task (AUC = .79) than the Conners' CPT (AUC = .62), and the combination of all three measures together resulted in the most sensitive and specific assessment with an AUC of .87 (Faraone et al., 2016). Throughout each of these studies, most analyses relied on collapsing multiple components (i.e., speed and accuracy) into larger general parameters; therefore, findings arguably lack precision and could not determine how RT, variability or skew in RT, omission, or failed inhibit errors might differentially contribute to this association.

To date, one study has used ROC analyses to examine the ability of ex-Gaussian parameters of RT to predict ADHD diagnosis. Studying boys in middle childhood using a 4-choice "Warned reaction time" task they found that tau and standard deviation were highly (both AUC = .96) diagnostic of ADHD, and though mean was also highly diagnostic (AUC = .90), mu and sigma held much lower diagnostic value (both AUC = .62) (Leth-Steensen et al., 2000). However, this study included a small sample size (N = 35), and only included boys. Therefore, while these findings strongly suggest that skewed reaction time data is both sensitive and specific in identifying school-aged boys with ADHD, further study with larger sample sizes, multiple age-ranges, and both boys and girls must be conducted to better understand the diagnostic

discriminability of skewed reaction time within the ADHD population.

Current Studies

This study used data collected from three archival ADHD datasets: (a) the "FRIENDS" study of 5-6-year-olds led by Drs. Karen Bierman and Cynthia Huang-Pollock; (b) the Child Attention and Learning "CAL" study of 8-12-year-olds led by Dr. Cynthia Huang-Pollock, and the (c) "Brain and Behavior" study of 18-25-year-olds led by Dr. Kristina Neely. Across all three datasets, ADHD diagnostic status was evaluated using a combination of clinical interview and age-normed standardized questionnaires; reaction time and inhibitory control was indexed using standard go/no-go and stop signal reaction time tasks. Specific methods are discussed individually for each study below.

The current project used Receiver Operating Characteristic (ROC) curves and the area under the curve (AUC) to examine the ability of computerized cognitive tasks to accurately identify individuals meeting diagnostic criteria for ADHD, then, develop possible clinical cut-offs for those variables found to have strong diagnostic discriminability. Additionally, by examining three different developmental periods, the current project examined how performance on these tasks, in relation to ADHD status, differs throughout development. The specific data analytic plan is discussed below.

Aims and Hypotheses

Aim 1

The first aim of this project is to assess the ability of reaction time and accuracy measures to discriminate among children and young adults with and without ADHD. Because previous analyses of sensitivity and specificity of ex-Gaussian predictors have been conducted exclusively

with boys (Leth-Steensen et al., 2000), analyses will be conducted separately by gender to evaluate if and how gender affects the ability of RT or accuracy to distinguish individuals with and without ADHD.

Hypotheses

1. If cognitive performance as indexed by SSRT, failed inhibit rate, mean RT, SDRT, sigma and tau can be used as sensitive and specific markers of ADHD, then area under the curve (AUC) values will be significantly higher than chance.
2. If cognitive impairment in individuals with ADHD is due in part to inconsistent/variable speeded performance, then SDRT and tau will discriminate between ADHD and non-ADHD participants comparably to standard indices of performance (i.e. SSRT and failed inhibit rate).

Aim 2:

The second aim of this project is to identify performance cutoffs that maximize the sensitivity and specificity in distinguishing between ADHD and typically development.

Methods

Study 1: FRIENDS Dataset

Participants. Boys and girls between the ages of 5-6 were recruited in two successive cohorts from 48 kindergarten classrooms in six Pennsylvania school districts that included both urban and rural areas, as part of a larger study examining the impact of a social-emotional intervention program on self-regulation.

Eligibility was determined in a two-step system. In the first step, kindergarten teachers completed two standardized behavioral rating scales; the Conner's ADHD Rating Scale – Revised, Short-Form (Conners, 2008) and the ADHD Rating Scale (DuPaul et al., 1998). For children with three or more inattentive or hyperactive/impulsive or four total symptoms endorsed by the teacher, parents then completed the Diagnostic Interview Schedule for Children version IV (DISC-IV) (Shaffer, Fisher, Lucas, Dulcan, & Schwab-Stone, 2000), as well as the Conner's Parent Rating Scale – Long-Form – Revised (Conners, 2008), and the Behavioral Assessment Scale for Preschool Children, 2nd edition (BASC-2; Reynolds & Kamphaus, 2004).

Children were determined to have ADHD (n = 75; 23 girls) if they met full clinical criteria for a diagnosis of ADHD on the DISC, including criteria for impairment, chronicity, and cross-situational severity. Additionally, at least one parent and one teacher report of behaviors on the Cognitive Problems/Inattention, Hyperactivity, ADHD Index, or DSM-IV Total Index of the Conner's Rating Scale or the Hyperactivity or Attention Problems Indices of the BASC was required to exceed the 85th percentile (T-score > 61), or at least three inattentive symptoms or hyperactive/impulsive symptoms or at least four total symptoms must have been endorsed as *often* or *very often* on the ADHD Rating Scale.

Typically developing children (n = 33, 15 girls) were recruited from the same classroom

and were classified as control participants if they did not meet diagnostic criteria for ADHD on the DISC-IV, teacher ratings of behavior on all relevant indices of the Conner's and BASC-2 were a T-score of 59 or below, and the total number of symptoms endorsed by parent or teacher (following the "or" algorithm; Lahey et al., 1994) yielded less than three inattentive or hyperactive/impulsive symptoms, and ≤ 3 total symptoms.

Children who did not meet full diagnostic criteria based on the DISC-IV, but received elevated parent or teacher ratings on behavioral questionnaires were considered to have "emerging" ADHD (n = 28; 12 girls), given that a lower number of symptoms are typically endorsed at these younger ages (Curchack-Lichtin, Chacko, & Halperin, 2014). Because the inclusion or exclusion of these children who exhibit symptoms but do not meet full diagnostic criteria is likely to influence the outcome of the study's primary analyses, primary analyses will be conducted both with and without these participants as part of a broader "non-ADHD" comparison group.

Children were excluded from the larger study if parents reported any sensorimotor disability, frank neurological disorder, or psychosis, if estimated Full Scale Intelligence Quotient (FSIQ) was below 70, if limited proficiency in English would impair full participation, or if legal custody was in dispute. Other common childhood disorders, such as anxiety, depression, oppositional defiant disorder, and conduct disorder were assessed using the DISC-IV, but was not exclusionary for either group.

Procedure. Parents gave informed written consent prior to participating. Data were collected by trained examiners at the children's school in a quiet room away from their classrooms. Estimated IQ was determined using the 2-subtest short form (Vocabulary, Matrices) of the Stanford Binet Fifth Edition (Roid, 2003).

Go-No-Go Task. Children completed a computerized visual Go-No-Go task, including four unique stimuli (blue triangle, blue square, red triangle, and red square). Each stimulus appeared for 1,000 ms; children were asked to make a key press when they saw a blue shape, but to withhold a response when they saw a red shape. They were given a total of 2,000 ms to respond, after which the next stimulus was presented regardless of whether a response had been made. If children responded within 2,000 ms, the next trial was immediately presented. Participants completed 60 consecutive trials; 70% were Go trials, 30% No-Go. Dependent variables were correct hits and failed inhibit errors, and mean and standard deviation of reaction time for all correct trials, and ex-Gaussian parameters of all correct trials. See Figure 4 for an illustration.

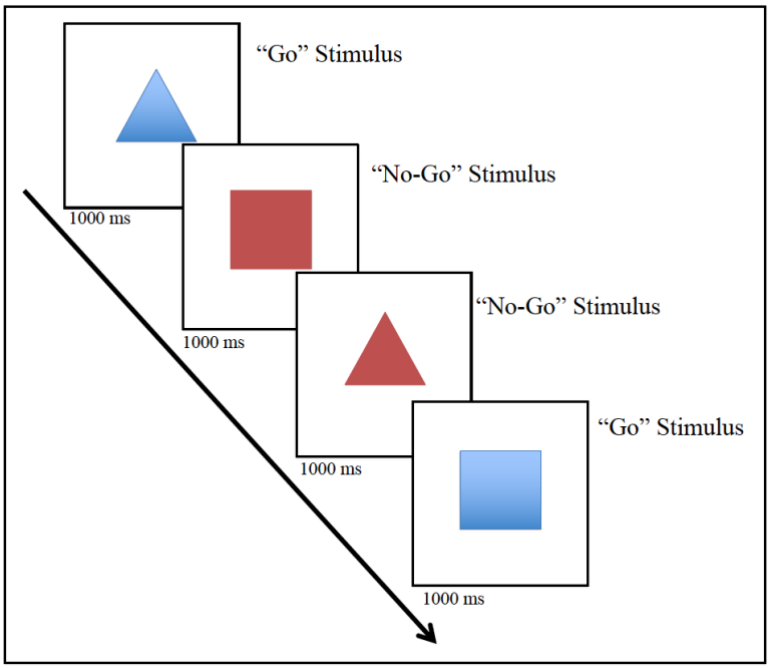

Figure 4. Illustration of the Go/No-Go task used in the FRIENDS project. Each stimulus appeared for 1,000 ms; children were asked to make a key press when they saw a blue shape, but to withhold a response when they saw a red shape. No feedback concerning accuracy was provided.

Study 2: CAL Dataset

The current study includes two separate but partially overlapping (N = 78) subsamples of participants from the larger CAL study; one subsample includes children who completed the Stop Signal Reaction Time (SSRT) task (total N = 292), the other subsample includes children who completed the Go/No-Go task (total N = 121), which was administered for only a short period of time during the multi-year CAL cross-sectional study. All procedures, including recruitment, exclusionary criteria, diagnostic decision-making, etc. are the same within each subsample and will only be described once. Demographic variables are described for all participants, and performance on each task is described separately.

Participants. Boys and girls between the ages of 8 and 12 years old were community recruited from Centre, York, and Harrisburg counties of Pennsylvania to participate in a larger study on attention and learning at The Pennsylvania State University. Children were excluded based on parent report of neurological or sensorimotor disorders, pervasive developmental disorder that would preclude full participation, and use of non-stimulant psychoactive medications (e.g., antidepressants) that could not be completely washed out before participation. Children with an estimated Full Scale IQ (FSIQ) < 80 were also excluded to prevent interpretation confounds.

Children were determined to have ADHD (N = 216, 73 girls) if they met DSM-IV criteria for ADHD including age of onset, duration, cross situational severity, and impairment as determined by parental report on the Diagnostic Interview Schedule for Children version IV (DISC-IV) (Shaffer et al., 2000). Additionally, at least one parent and one teacher report of behavior on the Attention, Hyperactivity, or ADHD subscales of the Behavioral Assessment Scale for Children (BASC-2; Reynolds & Kamphaus, 2004) or the Conners' Rating Scales

(Conners, 2008) was required to exceed the 85th percentile (T-score > 61). Both measures are commonly used and well-validated for the evaluation and diagnosis of ADHD. Following DSM-IV field trials, an "or" algorithm (Lahey et al., 1994) that integrated parent report on the DISC-IV and teacher report on the ADHD Rating Scale (DuPaul et al., 1998) was used to determine final symptom count. Children prescribed a psychostimulant medication (N = 83) were required to cease taking their medication a *minimum* of 24-48 hours in advance of the day of testing, though because many families opted to reduce the burden of medication washout by participating on weekends or during school breaks, the time since last dose varied (mean = 74.83, SD = 69.76).

Children were classified as control participants (N = 93; 46 girls) if they had never been diagnosed with or treated for ADHD in the past, parent and teacher reports on all above listed rating scales were below the 80th percentile (T-score ≤ 58), and 3 or fewer inattentive symptoms and 3 or fewer hyperactive/impulsive symptoms, and no more than 4 total symptoms using the "or" algorithm.

Children who did not meet full diagnostic criteria based on the DISC-IV, but whose parents/teachers endorsed 4 or 5 symptoms of inattention or hyperactivity/impulsivity using the "or" algorithm, were considered to have subthreshold symptomatology (N = 32, 17 girls). Because the inclusion or exclusion of these children who exhibit symptoms but do not meet full diagnostic criteria is likely to influence the outcome of the study's primary analyses, primary analyses will be conducted both with and without these participants.

The presence of common childhood disorders, such as anxiety, depression, oppositional defiant disorder, and conduct disorder was assessed using the DISC-IV, but the presence of these disorders was not exclusionary for either group.

Procedure. Informed written consent from parents and verbal assent from children was obtained prior to participating. Parents were compensated with a $100 gift card, provided with verbal clinical feedback on relevant test results, and children were allowed to choose a small toy (<$2) from a prize box. Participants completed the experimental paradigm as part of a larger battery across two separate testing sessions. A 2-subtest short form (Vocabulary, Matrix Reasoning) of the Wechsler Intelligence Scale for Children—Fourth Edition (WISC-IV: Wechsler, 2003) provided an estimated IQ. The correlation of the 2 subtest short form with the full 10-subtest battery is 0.87 (Sattler, 2008).

Stop Signal Reaction Time Task. Children were administered a 200-trial tracking version of the Logan Stopping Task (Logan, 1985). Children completed 20 practice trials to ensure understanding of the directions. The task was then administered in 5 blocks of 40 trials with an optional rest break between blocks. For each trial, a central fixation point appeared for 200 ms. On 75% of trials (i.e., "go" trials), an "X" or an "O" was presented in the center of the screen for 1000 ms, and children made a forced-choice response whether an "X" or an "O" had appeared. Children were given 2,300 ms to respond, after which the next trial automatically commenced. On 25% of trials (i.e., 50 "stop" trials), an auditory tone was presented to indicate that they should not respond. An initial mean reaction time (MRT) was determined based on the practice trials and the auditory stop tone was initially set to occur 250 ms before the MRT. The MRT was then dynamically recalculated after each correct go trial and the delay at which the stop tone was presented was adjusted dynamically in 50 ms increments to maintain an overall ~50% accuracy rate. Stop signal reaction time (SSRT), the amount of time a child needs to successfully inhibit a response 50% of the time, was calculated by subtracting the mean stop signal delay from the child's MRT. See Figure 5 for illustration.

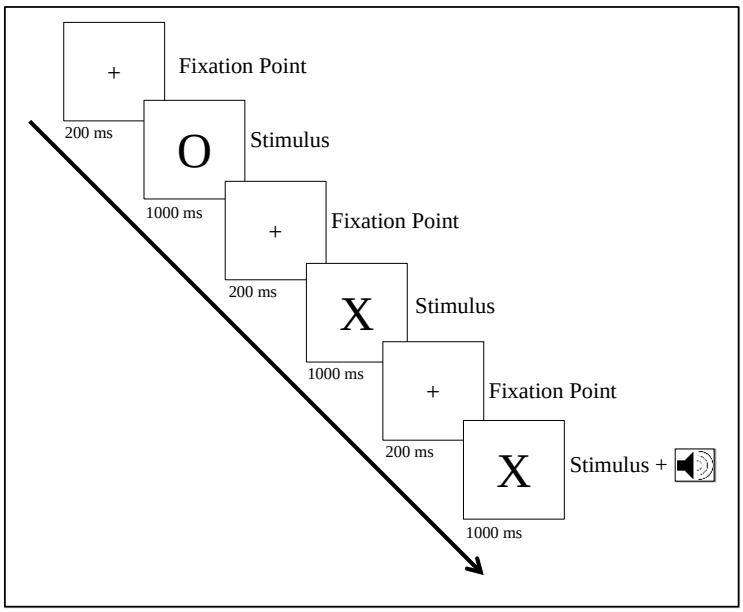

Figure 5. Illustration of the Stop Signal Reaction Time (SSRT) task used in the CAL Study. When an "X" or "O" stimulus was displayed, participants were to respond by pressing the correct "X" or "O" key on a keyboard. If a computerized tone was played during stimulus presentation, participants were to refrain from pressing any keys. No feedback concerning accuracy was provided.

Go-No-Go Task. The GNG paradigm used stimuli from a numerosity discrimination task (Gomez, Ratcliff, & Perea, 2007; Ratcliff, Love, Thompson, & Opfer, 2012). Blocks of 80 trials were administered with optional rest periods in between. At the start of each trial, a number of asterisks filled random positions in a 10×10 white box centered on a black screen. Children were told "We're going to play a game called the Candy Factory now. Some of the boxes of candy that the factory makes have a lot of candy in them, and some only have a little. But, the sorter is broken! We need your help! Every time you see a box that has 'a lot' of candy, press the spacebar. Don't press anything if the box has "a little" bit of candy. This is a hard game, but try to work as quickly as you can without making mistakes. Let's try some for practice." Four practice trials were completed to ensure comprehension of instructions. 75% of the stimuli (selected at random without replacement) were "go" trials and contained 61–70 asterisks. The

remaining “no-go” trials contained 31–40 asterisks. Children were not provided any instruction on how to distinguish “a lot” from “a little”, but were provided with a brief error tone after incorrect responses. Stimuli were presented for a maximum of 1500 ms, with the next trial beginning 300 ms after a response. See Figure 6 for illustration. Dependent variables were correct hits and failed inhibit errors, and mean and standard deviation of reaction time for all correct trials, and ex-Gaussian parameters of all correct trials. This task was completed at the second testing session (following the exclusion of children with subthreshold symptomatology) and for a short period of time; as such, only a relatively small subset of participants completed this task. AUC values for this task therefore will not be directly compared to the SSRT task.

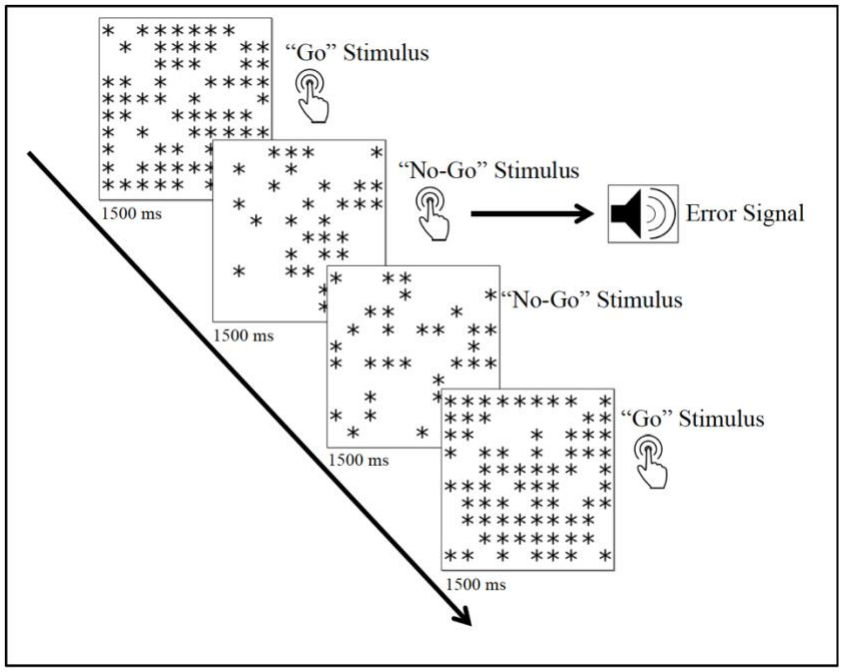

Figure 6. Illustration of the Go/No-Go task used in the CAL Study. Children were told to press a key if there were “a lot” of asterisks, and not press a key if there were only “a few” asterisks. They were not provided any instruction on how to distinguish “a lot” from “a few”, but were provided with a brief error tone after incorrect responses. The figure depicts a correct “go” trial, an incorrect “no-go” trial in which they failed to inhibit a response and received an error signal, a correct “no-go” trial in which they inhibit a response, and a correct “go” trial.

Study 3: Brain and Behavior Dataset

Participants. Young adults aged 18–25, with and without ADHD, were recruited from Centre County in Pennsylvania to participate in a larger study on attention and motor coordination at The Pennsylvania State University. Potential participants were excluded based on report of previous concussions that resulted in a loss of consciousness for more than 10 min; previous diagnosis of seizures, epilepsy, encephalitis, meningitis or an autism spectrum disorder; previous diagnosis of a musculoskeletal or neurological disorder; or previous diagnosis of any disorder involving psychosis. Participants with an estimated Full Scale IQ (FSIQ) <80 were also excluded to prevent interpretation confounds.

Adults were determined to have ADHD (N = 62, 35 women) if they met DSM-V criteria for ADHD including at least 5 symptoms of inattention or hyperactivity/impulsivity, and cross situational severity and impairment as determined by self-report on: the Connors Adult ADHD Rating Scales (CAARS), the Achenbach Adult Self Report (ASR) (Achenbach, 2003), and the Conners' Adult ADHD Diagnostic Interview (CAADID); a semi-structured interview (Conners, Epstein, & Johnson, 2001). Participants prescribed a psychostimulant medication (N = 42) were not required to cease taking their medication in advance of the day of testing; therefore, unlike in either the FRIENDS or CAL samples, performance on this Go/No-Go task reflects their best possible performance when treating their ADHD symptoms rather than best possible performance without the potential benefit of treatment. No participants were taking antipsychotics, or anticonvulsants.

Adults were classified as control participants (N = 72, 30 women) if they had never been diagnosed with or treated for ADHD in the past, and reported fewer than 2 symptoms of inattention or hyperactivity/impulsivity on the CAADID.

Adults who did not meet full diagnostic criteria but who endorsed 3 or 4 symptoms of inattention or hyperactivity/impulsivity on the CAADID, were considered to have subthreshold symptomatology (N = 33, 16 women). Because the inclusion or exclusion of these individuals who exhibit symptoms but do not meet full diagnostic criteria is likely to influence the outcome of the study's primary analyses, primarily analyses will be conducted both with and without these participants.

The presence of anxiety or depression and other common concerns was also assessed by self-report using the Adult Self-Report Scale (Achenbach & Rescorla, 2003) but was not exclusionary for either group.

Procedure. Informed written consent was obtained prior to participating. Participants completed the Connors Adult ADHD Rating Scales (CAARS) and the Achenbach Adult Self Report (ASR) (Achenbach, 2003) as well as more general self-reports of mental health and daily functioning before attending the in-person laboratory session. Participants completed the experimental paradigm as part of a larger battery of cognitive tasks. A 2-subtest short form (Vocabulary, Matrix Reasoning) of the Wechsler Adult Intelligence Scale- Fourth Edition (WAIS-IV; Wechsler, 2008) provided an estimated IQ. Participants were compensated with $40 following the laboratory session.

Go-No Go Task. Participants completed a computerized visual Go-No-Go task, including six unique stimuli; colored horizontal bars (red, green, aqua, orange, yellow, or blue), presented on a black background. Trials were 750 ms in duration, followed by a 500 ms inter-trial period of rest. Each participant completed four blocks of 100 trials (25% no-go) for a total of 400 trials. Each block started and ended with 12.5 seconds of rest. Go and No-Go trials were randomly presented. Participants were instructed to make a key press as quickly as possible ("go" trials) if

the bar was colored green, aqua, orange, or yellow; they were instructed not to make a key press ("no-go" trials) if the bar was blue. See Figure 7 for Illustration. Each participant practiced the task immediately prior to the experimental session. No feedback concerning accuracy was provided. Dependent variables were correct hits and failed inhibit errors, and mean and standard deviation of reaction time for all correct trials, and ex-Gaussian parameters of all correct trials.

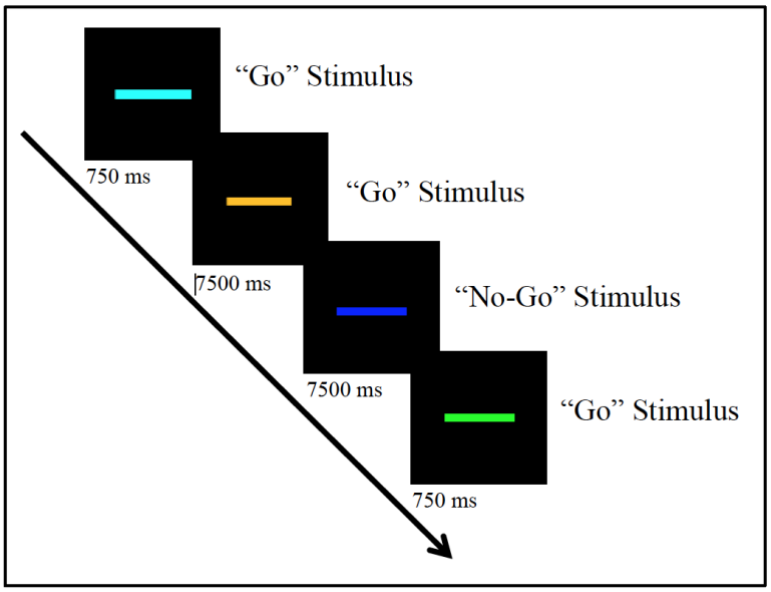

Figure 7. Illustration of the Go/No-Go task used in the Brain and Behavior Study. Red, orange, green, and aqua bars were the "go" signal instructing participants to respond with a keypress; Blue bars were the "no-go" signal instructing participants not to respond. No feedback concerning accuracy was provided.

Data Analysis Plan

Sample Demographics

Initial analyses (i.e., ANOVA) were conducted to compare ADHD and non-ADHD groups on key demographic variables (e.g., age, IQ, gender) to ensure equivalence of groups and limit possible misinterpretation due to confounding variables.

Dependent variables

Stop Signal Reaction Time (SSRT). A common measure of inhibitory control, SSRT reflects the amount of time needed to successfully inhibit a response 50% of the time. It is calculated by subtracting the mean stop signal delay from mean RT.

Go-No-Go Task % Failed inhibits. For Go/No-Go tasks, accuracy are defined as the percent of correct "hits" (i.e.., correctly responding with a keypress to a "go" trial), for all trials excluding practice blocks. Errors will therefore be measured as the percent of failed inhibits, or, the percent of "go" keypresses on "no-go" trials.

Reaction time (RT) distributions. For both the SSRT and GNG tasks, mean and standard deviation of reaction time were calculated using all correct "go" responses, not preceded by a "no-go" trial, so as to reduce the possible effect of inhibitory control on the RT distribution (Schachar et al., 2004). Additionally, trials < 150 milliseconds were excluded as they are generally thought to represent anticipatory error.

The normality of reaction time distributions for each task were assessed using traditional data exploration techniques. Ex-Gaussian parameters mu, sigma, and tau will be computed using an egfit tool in MATLAB (Lacouture & Cousineau, 2008), using the same RT trial criteria described for mean and standard deviation. This function performs an iterative search process to compare the observed RT distribution to an ex-Gaussian probability density function (PDF) using a Simplex method. In each iteration, the parameter values of the PDF are adjusted until maximum fit to the observed data is achieved. To validate that the ex-Gaussian distribution is an appropriate fit for the data, a simulation-recovery study was conducted using the "rexgauss()" function from the R package "retimes." This generates simulated response time data based on the ex-Gaussian parameters derived from the empirical response times, which is then fit to the ex-Gaussian distribution using the same procedures described above. Correlations greater than or

equal to .80, indicate the data fits well.

Hypothesis Testing

Aim 1: Sensitivity and Specificity. ANOVAs will be used to determine diagnostic group differences in performance.

Hypothesis 1. To test the hypothesis that cognitive performance as indexed by SSRT, failed inhibit rate, mean RT, SDRT, sigma and tau can be used as sensitive and specific markers of ADHD, ROC curves and AUC statistics will be derived using standard procedures in SPSS version 25; significance will indicate if the measure discriminates better than chance.

Hypotheses 2. To test the hypotheses that variable reaction time can function as a marker of ADHD as well as traditional measures such as SSRT and failed inhibition, and that ex-Gaussian parameters more effectively capture variable reaction time throughout a task, AUC values, their standard errors, and the correlations among predictors in both groups will be used to compare the discriminability (sensitivity and specificity) of each variable within each sample. If this is true, it is expected that SDRT and tau will be at least as sensitive and specific as SSRT and % failed inhibits, and tau will be more sensitive and specific than SDRT. It is anticipated that mean RT, mu, and sigma will not be as sensitive or specific as any of the other variables.

Post-Hoc Analyses. For any accuracy or RT variable with significant AUC values, the above analyses will also be conducted separately by gender to explore if accuracy or variability in RT differentially distinguish between individuals with and without ADHD based on gender.

Aim 2: Creating Cutoffs. For each parameter found to distinguish between ADHD and non-ADHD participants, exploratory analyses will be conducted to determine optimal cutoffs that maximize both sensitivity and specificity to consider if and how these measures might be most usefully applied in practice.

Results

Results will be discussed thematically and by aim/hypotheses rather than separately by sample to best highlight the ways in which findings are consistent/inconsistent across samples.

Validating ex-Gaussian Parameters.

Within each dataset, we conducted a simulation-recovery study to validate the ex-Gaussian distribution as an appropriate fit for the response time distributions. Using the "rexgauss()" function from the R package "retimes," we generated simulated response time data based on the ex-Gaussian parameters derived from the empirical response times. These simulated RT trials were then fit to the ex-Gaussian distribution using the same procedures described above. For the CAL SSRT, CAL GNG, and Brain and Behavior (B&B) GNG tasks, correlations between the simulated and empirical values for mu, sigma, and tau were all above 0.8, indicating the parameters fit well. For the FRIENDS sample, the correlation between simulated and empirical mu was above .8. Sigma and tau were both low (.4), indicating poor fit, likely due to the low number of trials within the task impeding accuracy of data simulation and validation, rather than the distribution of reaction times not fitting the ex-Gaussian distribution. Therefore, analyses proceeded with FRIENDS using the ex-Gaussian parameters, though the discussion section will address limitations associated with uncertain fit.

Diagnostic Group Differences.

Tables 1-3 provide descriptive statistics for demographic and diagnostic information within each sample. Validating the diagnostic groups, there were significant differences in the number of inattentive and hyperactive symptoms across groups (all $p < .001$, and $\eta 2 > .2$). There were no group differences in age (all $F < .82$, $p > .37$, $\eta 2 < .001$). Consistent with expectations, there were more males than females with ADHD, compared to typically developing controls.

There was a significant group difference in FSIQ in the FRIENDS and B&B samples (ranging from 98-112; all $p < .02$, all $\eta 2 > .04$), however, they are well within the population average.

	Control	***ADHD***	**Test Statistics**	***p***
N (Boys:Girls)	33 (18:15)	75 (52:23)	$\chi 2$ (1 df) = 219	$p = .104$
Age in years	5.30 (.47)	5.27 (.45)	F(1, 106) = .15 $\eta 2 = .001$	$p = 0.701$
Estimated FSIQ	107.27 (8.94)	98.56 (12.86)	F(1, 106) = 12.46 $\eta 2 = .105$	$p = .001$**
Inattention				
Total # of Symptoms	0.30 (.81)	6.49 (2.12)	F(1, 106) = 2.63.16 $\eta 2 = .713$	$p < .001$**
Parent BASC-2	43.55 (6.81)	60.60 (8.50)	F(1, 106) = 103.38 $\eta 2 = .494$	$p < .001$**
Parent Conners	43.51 (3.56)	60.77 (11.38)	F(1, 106) = 72.28 $\eta 2 = .408$	$p < .001$**
Teacher Conners	46.64 (3.16)	64.47 (16.54)	F(1, 106) = 37.55 $\eta 2 = .262$	$p < .001$**
Hyperactivity/Impulsivity				
Total # of Symptoms	0.39 (.83)	6.87 (2.44)	F(1, 106) = 220.13 $\eta 2 = .675$	$p < .001$**
Parent BASC-2	45.94 (5.63)	63.57 (10.66)	F(1, 106) = 80.12 $\eta 2 = .430$	$p < .001$**
Parent Conners	46.61 (5.01)	66.03 (11.13)	F(1, 106) = 91.82 $\eta 2 = .464$	$p < .001$**
Teacher Conners	44.21 (1.53)	67.79 (13.27)	F(1, 106) = 102.04 $\eta 2 = .490$	$p < .001$**

**=p<.01; *=p<.05

Table 1. Description of groups in the FRIENDS sample. Means, with standard deviation in parentheses. All ratings scales reported in T-scores unless otherwise noted.

	Control	***ADHD***	***Test Statistics***	***P***
N (Boys:Girls)	93 (47:46)	216 (143:73)	$\chi 2$ (1 df) = 6.82	p = .007**
Age in years	9.66 (1.30)	9.50 (1.21)	F(1,293) = .820 $\eta 2$ = .003	p=.366
Estimated FSIQ	105.01 (8.54)	103.29 (12.81)	F(1, 293) = .800 $\eta 2$ = .003	p = .372
Inattention				
Total # of Symptoms	.50 (.65)	7.94 (1.59)	*F(1, 293) = 1959.50 $\eta 2$ = .870*	p < .001**
Parent BASC-2	43.54 (6.488	66.78 (6.91)	*F(1, 293) = 717.13 $\eta 2$ = .710*	p < .001**
Teacher BASC-2	43.22 (5.43)	61.66 (7.05)	*F(1, 293) = 466.502 $\eta 2$ = .616*	p < .001**
Parent Conners	46.02 (4.08	70.75 (11.37)	*F(1, 293) = 389.69 $\eta 2$ = .571*	p < .001**
Teacher Conners	46.20 (4.30)	59.88 (11.34)	*F(1, 293) = 114.663 $\eta 2$ = .281*	p < .001**
Hyperactivity/Impulsivity				
Total # of Symptoms	.23 (.49)	5.70 (2.74)	*F(1, 293) = 352.760 $\eta 2$ = .547*	p < .001**
Parent BASC-2	42.31 (5.20)	65.94 (13.50)	*F(1, 293) = 254.169 $\eta 2$ = .465*	p < .001**
Teacher BASC-2	43.85 (3.85)	60.22 (12.51)	*F(1, 293) = 140.298 $\eta 2$ = .325*	p < .001**
Parent Conners	45.89 (3.62)	68.91 (14.26)	*F(1, 293) = 223.743 $\eta 2$ = .433*	p < .001**
Teacher Conners	45.31 (2.96)	59.04 (11.88)	F(1, 293) = 115.737 $\eta 2$ = .283	p < .001**
Comorbidity (DISC-IV)				
MDD/Dysthymia	0/0	10/4	$\chi 2$ (1 df) = 5.344	p = 0.013*
GAD	1	26	$\chi 2$ (1 df) = 9.25	p = .001**
ODD/CD	3/0	81/20	$\chi 2$ (1 df) = 38.96	p < .001**
Go/No-Go Task				
N (Boys:Girls)	35 (19:16)	70 (46:24)	$\chi 2$ (1 df) = 1.292	p = .178
SSRT Task				
N (Boys:Girls)	76 (34:42)	194 (128: 66)	$\chi 2$ (1 df) = 10.27	p = .001**

**=p<.01; *=p<.05

Table 2. Description of groups in the CAL sample. Gender, age, IQ, symptoms of ADHD, and rates of comorbidity are described for the entire sample. GNG and SSRT task variables are described only for the subsamples which completed each task. Means, with standard deviation in parentheses. BASC and Conners reported in T-scores. Comorbidity variables are the number of children meeting diagnostic criteria for each disorder based on parent DISC.

	Control	*ADHD*	*Test Statistics*	*P*
N (Men:Women)	72 (42:30)	61 (26:35)	$\chi 2$ (1 df) = 3.26	p = 0.051
Age in years	21.19 (.50)	21.13 (1.80)	$F(1, 131) = .035$ $\eta 2 < .001$	p = .851
Estimated FSIQ	110.32 (11.73)	105.25 (11.08)	$F(1, 131) = 6.499$ $\eta 2 = .047$	p = .012*
Inattention				
Total # of Symptoms	.04 (.26)	6.21 (2.05)	*F(1, 131) = 640.835 η2 = .830*	p < 0.001**
CAARS Inattention	46.21 (9.46)	67.39 (15.43)	*F(1, 131) = 247.873 η2 = .654*	p < 0.001**
Hyperactivity/Impulsivity				
Total # of Symptoms	.10 (.34)	4.61 (2.40)	*F(1, 131) = 94.092 η2 = .418*	p < 0.001**
CAARS Hyperactivity	42.25 (8.54)	57.49 (13.76)	$F(1, 131) = 60.765$ $\eta 2 = .317$	p < 0.001**
Comorbidity				
Anxious/Depressed	51.56 (3.38)	59.92 (10.80)	*F(1, 131) = 24.015 η2 = .155*	p < 0.001**
Aggressive Behavior	44.58 (10.43)	56.49 (7.45)	*F(1, 131) = 25.430 η2 = .163*	p < 0.001**
Internalizing	44.58 (10.43)	54.90 (13.04)	*F(1, 131) = 25.692 η2 = .164*	p < 0.001**
Externalizing	44.54 (9.91)	56.10 (10.22)	$F(1, 131) = 43.647$ $\eta 2 = .250$	p < 0.001**

**=p<.01; *=p<.05

Table 3. Description of groups in the Brain and Behavior sample. Means, with standard deviation in parentheses. BASC and Conners reported in T-scores. Comorbidity variables are the number of children meeting diagnostic criteria for each disorder based on parent DISC.

Table 4 provides descriptive statistics for cognitive task variables across all samples. Diagnostic group differences were found such that individuals with ADHD had significantly greater failed inhibit error rate in the GNG tasks in all samples, and longer SSRT in the CAL SSRT task (all $p < .05$, all $\eta 2 > .046$). Of the RT distribution variables, individuals with ADHD had significantly longer SDRT in all tasks and all samples (all $p < .01$, all $\eta 2 > .048$), and longer tau in all tasks except FRIENDS GNG (FRIENDS $p = .215$, $\eta 2 = .020$; CAL and B&B all $p < 05$, all $\eta 2 > .017$).

	FRIENDS			CAL SSRT			CAL GNG			B&B GNG		
	Control	***ADHD***	**Test Statistics**	***Control***	***ADHD***	**Test Statistics**	***Control***	***ADHD***	**Test Statistics**	***Control***	***ADHD***	**Test Statistics**
SSRT	--	--	--	333.48 (108.23)	441.57 (150.31)	$F(1,269) = 32.64^{**}$ $\eta2 = .109$	--	--	--	--	--	--
% Failed Inhibits	15.32 (10.30)	23.04 (18.37)	$F(1,106) = 5.10^{*}$ $\eta2 = .046$	--	--	--	31.60 (14.03)	43.49 (15.90)	$F(1,104) = 14.06^{**}$ $\eta2 = .120$	11.21 (8.55)	21.13 (14.64)	$F(1,132) = 23.61^{**}$ $\eta2 = .153$
RT Variables												
mRT	660.54 (125.83)	711.87 (143.92)	$F(1,106) = 3.14$ $\eta2 = .029$	791.50 (167.20)	821.46 (169.59)	$F(1,269) = 1.72$ $\eta2 = .006$	583.59 (85.91)	602.23 (87.37)	$F(1,104) = 1.07$ $\eta2 = .01$	355.85 (37.81)	348.01 (47.21)	$F(1,132) = 1.13$ $\eta2 = .009$
SDRT	199.77 (63.11)	242.93 (71.84)	$F(1,106) = 8.88^{**}$ $\eta2 = .077$	209.49 (167.20)	247.95 (80.78)	$F(1,269) = 13.42^{**}$ $\eta2 = .048$	187.51 (35.46)	219.58 (38.75)	$F(1,104) = 16.90^{**}$ $\eta2 = .141$	64.50 (12.51)	73.38 (15.82)	$F(1,132) = 13.07^{**}$ $\eta2 = .091$
RT mu	486.22 (137.47)	510.11 (170.65)	$F(1,106) = .50$ $\eta2 = .005$	621.84 (157.35)	624.07 (181.56)	$F(1,269) = .009$ $\eta2 < .001$	412.81 (68.96)	401.99 (89.07)	$F(1,104) = .40$ $\eta2 = .004$	299.74 (35.46)	282.52 (41.44)	$F(1,132) = 6.67^{*}$ $\eta2 = .048$
RT sigma	83.66 (62.32)	108.98 (77.95)	$F(1,106) = 2.71$ $\eta2 = .025$	111.83 (47.98)	134.12 (72.61)	$F(1,269) = 6.11^{*}$ $\eta2 = .022$	76.21 (27.48)	93.32 (35.58)	$F(1,104) = 6.23^{*}$ $\eta2 = .057$	29.52 (11.21)	31.64 (11.65)	$F(1,132) = 1.14$ $\eta2 = .009$
RT tau	174.30 (74.89)	202.04 (96.63)	$F(1,106) = 2.15$ $\eta2 = .020$	169.66 (96.93)	197.40 (93.78)	$F(1,269) = 4.69^{*}$ $\eta2 = .017$	170.78 (41.31)	200.24 (47.57)	$F(1,104) = 9.74^{**}$ $\eta2 = .086$	56.12 (14.48)	65.49 (16.53)	$F(1,132) = 12.17^{**}$ $\eta2 = .085$

Table 4. ADHD vs Control diagnostic group differences in GNG and SSRT task variables, across all samples. Means, with standard deviation in parentheses.

Aim 1: Sensitivity and Specificity.

Determining Discriminability. The diagnostic utility of each task parameter was assessed using ROC curves and evaluating the AUC. Results will be discussed across all samples, noting consistent and inconsistent findings; Table 5 provides the AUC statistics, and Figure 8 shows ROC graphs for each sample. As expected given their traditional use in evaluations of ADHD performance on these tasks, SSRT discriminated between ADHD and typically developing participants at a level better than chance (AUC = .717, $p < .001$) in the CAL sample. Percent failed inhibits similarly discriminated between groups in the CAL and B&B GNG tasks (both AUC > .708, both $p < .001$) but not FRIENDS (AUC = .608, $p = .076$). Of the reaction time distribution variables, SDRT discriminated between groups in all tasks, (all AUC > .659, all $p < .005$). When the distribution was decomposed into ex-Gaussian parameters, the differences in SDRT appear to be driven primarily by tau (AUC > .593, $p < .02$) for both the school-aged (CAL) and young adult (B&B) samples. Sigma was only significant in the middle-childhood CAL sample (AUC > .592, $p < .02$). Decompisiton of the RT distribution did not provide any additional information for the kindergarten-aged FRIENDS sample.

		FRIENDS GNG			CAL SSRT			CAL GNG			B&B GNG		
		AUC	Std. Error	Asymp. Sig	AUC	Std. Error	Asymp. Sig	AUC	Std. Error	Asymp. Sig	AUC	Std. Error	Asymp. Sig
Boys and Girls	SSRT	--	--	--	0.717	0.033	<.001**	--	--	--	--	--	--
	% Failed Inhibits	0.608	0.054	0.076	--	--	--	0.708	0.053	0.001**	0.730	0.044	<.001**
	mean RT	0.605	0.058	0.084	0.569	0.04	0.078	0.556	0.060	0.352	0.418	0.050	0.103
	SDRT	0.671	0.055	0.005**	0.659	0.037	<.001**	0.727	0.051	<.001**	0.666	0.047	0.001**
	RT mu	0.543	0.060	0.482	0.498	0.041	0.956	0.432	0.059	0.256	0.334	0.048	0.001**
	RT sigma	0.581	0.060	0.179	0.592	0.039	0.018**	0.694	0.056	0.001**	0.544	0.051	0.383
	RT tau	0.606	0.056	0.079	0.593	0.038	0.018**	0.691	0.055	0.002**	0.655	0.047	0.002**
Boys Only	SSRT	--	--	--	0.755	0.041	<.001**	--	--	--	--	--	--
	% Failed Inhibits	0.494	0.068	0.941	--	--	--	0.692	0.073	0.016**	0.728	0.060	0.002**
	mean RT	0.744	0.070	0.002**	0.551	0.055	0.361	0.580	0.082	0.313	0.464	0.072	0.623
	sd RT	0.728	0.068	0.004**	0.711	0.049	<.001**	0.720	0.069	0.006*	0.700	0.066	0.006**
	RT mu	0.661	0.076	0.042*	0.445	0.055	0.324	0.442	0.082	0.462	0.380	0.071	0.098
	RT sigma	0.638	0.079	0.083	0.536	0.056	0.521	0.672	0.078	0.031*	0.616	0.074	0.109
	RT tau	0.641	0.068	0.076	0.666	0.052	0.003*	0.727	0.072	0.004**	0.693	0.066	0.008**
Girls Only	SSRT	--	--	--	0.688	0.053	0.001**	--	--	--	--	--	--
	% Failed Inhibits	0.745	0.080	0.012**	--	--	--	0.678	0.085	0.059	0.732	0.063	0.001**
	mean RT	0.513	0.095	0.893	0.617	0.058	0.041*	0.576	0.093	0.423	0.389	0.071	0.124
	sd RT	0.632	0.093	0.174	0.664	0.054	0.004**	0.758	0.078	0.006**	0.630	0.069	0.074
	RT mu	0.420	0.092	0.411	0.55	0.06	0.381	0.456	0.093	0.639	0.301	0.065	0.006**
	RT sigma	0.520	0.098	0.834	0.636	0.055	0.017*	0.727	0.083	0.016*	0.476	0.073	0.742
	RT tau	0.568	0.097	0.483	0.587	0.056	0.129	0.698	0.084	0.036*	0.610	0.071	0.130

Table 5. Area under the curve statistics for all tasks: FRIENDS GNG, CAL SSRT, CAL GNG, and B&B GNG. Distinguishing between individuals with ADHD and typically developing peers.

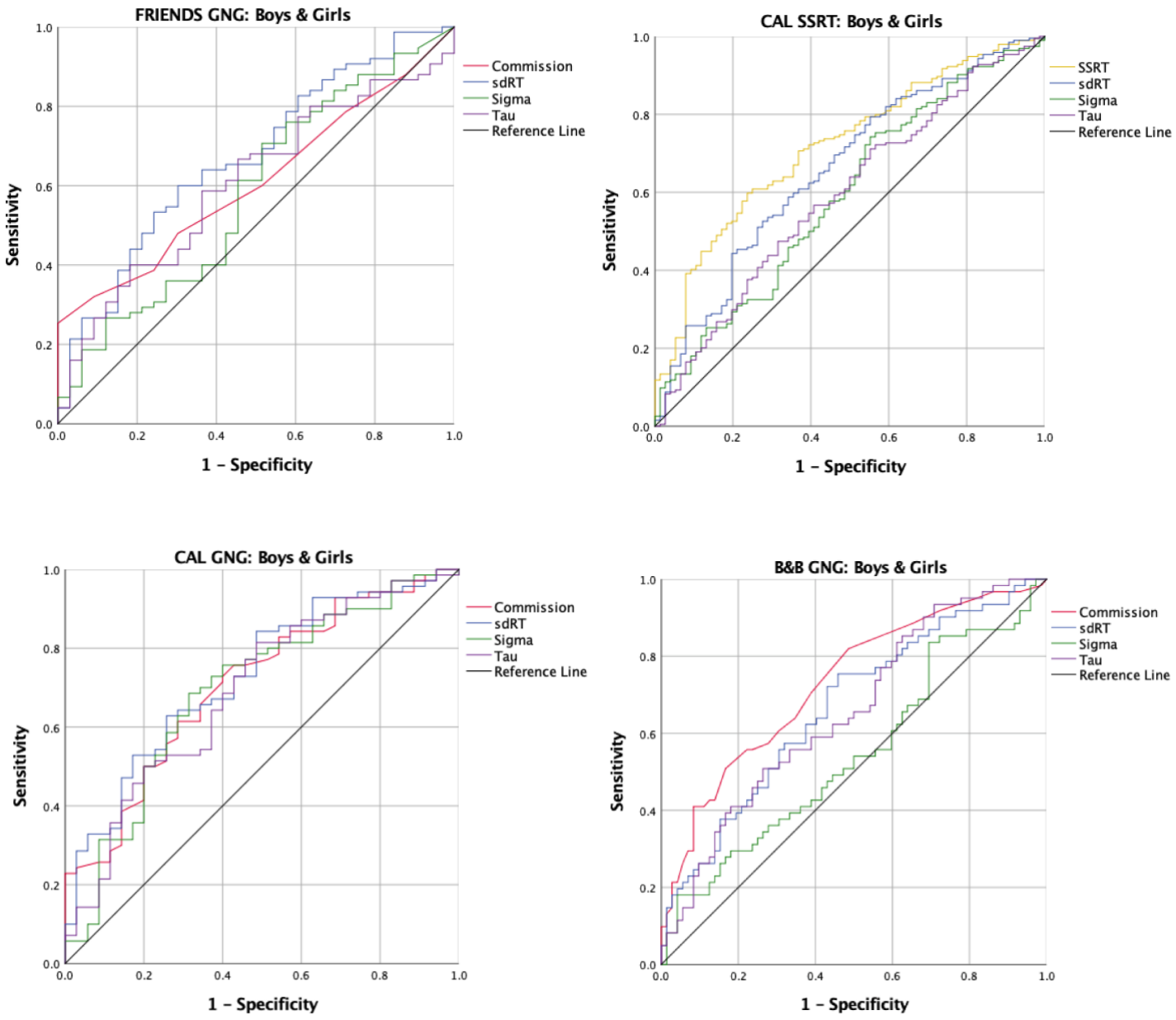

Figure 8. ROC curves for all tasks, distinguishing children with ADHD from typically developing peers, comparing boys and girls together.

Gender differences. When post-hoc analyses were completed to determine if discriminability varied by gender, results were consistent for boys (See Table 5). However, for girls, the primary change was that SDRT was only discriminatory in the two CAL tasks (AUC >.664, $p < .004$), and that affect appeared to be driven by both sigma (AUC > .636, $p < .017$) and tau (AUC = .698, $p = .036$).

Comparing Strength of Predictive Utility. To statistically compare the diagnostic discriminability of

SSRT, failed inhibit, and RT variables, Z-scores representing differences in the strength of discriminability between two variables were generated. Z-scores were created by using their AUC values, standard errors, and the correlations between the two predictors within each group (Hanley & McNeil, 1983) (See Table 6). For the school-aged CAL sample, both SSRT and SDRT predicted ADHD status better than tau. All other comparisons of performance variables were equally predictive of ADHD, with none being stronger than the other (all $z < 1.38$, all $p > .18$).

		FRIENDS GNG				CAL SSRT				CAL GNG				B&B GNG			
		r Control	r ADHD	z	*p*	r Control	r ADHD	z	*p*	r Control	r ADHD	z	*p*	r Control	r ADHD	z	*p*
Boys and Girls	SSRT vs SDRT	--	--	--	--	0.21	0.337	1.34	0.18	--	--	--	--	--	--	--	--
	SSRT vs tau	--	--	--	--	0.258	0.476	3.03	0.003**	--	--	--	--	--	--	--	--
	FI vs SDRT	-0.028	0.217	-0.85	0.40	--	--	--	--	0.113	0.212	-0.28	0.78	0.105	-0.109	0.104	0.30
	FI vs tau	0.118	0.209	+	+	--	--	--	--	0.041	0.127	0.23	0.82	0.008	-0.092	0.119	0.23
	SDRT vs tau	0.721	0.616	1.38	0.17	0.836	0.65	2.37	0.018**	0.929	0.804	1.19	0.23	0.731	0.937	0.37	0.71
Boys Only	SSRT vs SDRT	--	--	--	--	0.142	0.337	n0.77	0.442	--	--	--	--	--	--	--	--
	SSRT vs tau	--	--	--	--	0.193	0.53	1.63	0.103	--	--	--	--	--	--	--	--
	FI vs SDRT	0.142	0.354	+	+	--	--	--	--	0.198	0.208	-0.31	0.76	0.069	-0.22	0.33	0.74
	FI vs tau	0.269	0.232	+	+	--	--	--	--	0.153	0.132	-0.31	0.71	-0.116	-0.1	0.41	0.68
	SDRT vs tau	0.687	0.733	1.6	0.11	0.732	0.533	0.98	0.326	0.914	0.956	-0.2	0.84	0.561	0.944	0.14	0.89
Girls Only	SSRT vs SDRT	--	--	--	--	0.206	0.373	0.37	0.712	--	--	--	--	--	--	--	--
	SSRT vs tau	--	--	--	--			^		--	--	--	--	--	--	--	--
	FI vs SDRT			^		--	--	--	--	0.062	0.369	-0.79	0.43	0.174	-0.15	1.18	0.24
	FI vs tau			^		--	--	--	--	-0.022	0.168	-0.17	0.86			^	
	SDRT vs tau			^				^		0.946	0.482	0.9	0.37			^	

^ AUC comparisons were not tested if either variable was not significantly discriminatory.

+ Despite significance, the average of the two AUCs are too low to reliably compare (Hanley & McNeil, 1983).

Table 6. Comparison of AUC values across all tasks, distinguishing between individuals with ADHD and typically developing peers. (FI= "percent failed inhibits").

Gender differences. No comparisons indicated that any variables were stronger than another when examining boys and girls separately.

Aim 1 Summary. Overall, results indicated that impaired performance on cognitive tasks predicts ADHD diagnosis above chance. As expected, this was true for SSRT and failed inhibit rate, the standard (traditional) variables evaluated in these tasks. SDRT also predicted ADHD diagnosis better than chance, and examination of the ex-Gaussian parameters of the RT distribution indicated that this is largely driven by tau, though not in the younger kindergarten-age sample. These results held for boys, but were somewhat less evident in girls. Though the size of values dropped somewhat when individuals with subclinical symptomatology were included in analyses, the general patterns of results did not.

Aim 2: Creating Cutoffs.

Following the identification of variables that were able to accurately discriminate between individuals with and without ADHD, the sensitivity and specificity of each was evaluated to determine optimal cutoffs in the determination of ADHD. When applied to the performance of an individual, these cutoffs would represent the point at which performance would be considered indicative of ADHD. No consensus exists regarding a single preferred method to determine optimal cutoffs (Hajian-Tilaki, 2018), and the method chosen for a study is often based on a priori conceptualizations of what scores should be considered "elevated", or prioritization of either sensitivity or specificity given the particular aims of the study (Henry, Grant, & Cropsey, 2018; Hutchinson, Bradbury, Browne, & Hurley, 2017). However, one well-established method is the use of Youden's Index, in which a single parameter is used to determine the point at which the combined accuracy of sensitivity and specificity are highest (Youden's Index = Sensitivity + Specificity-1), without prioritizing one over the other (Yin &

Tian, 2014; Youden, 1950). This therefore makes it an optimal method when there are not strong arguments for valuing sensitivity over specificity or vice versa. Table 7 includes the optimal cutoff scores, sensitivity, specificity, and Youden's Index for SSRT, failed inhibits, SDRT, and tau in all tasks and samples.

		FRIENDS GNG				CAL SSRT				CAL GNG				B&B GNG			
		Value	Sens.	Spec.	Youden	Value	Sens.	Spec.	Youden	Value	Sens.	Spec.	Youden	Value	Sens.	Spec.	Youden
Boys and Girls	SSRT	--	--	--	--	445.89	0.47	0.85	0.32	--	--	--	--	--	--	--	--
	% Failed Inhibits	36.11	0.25	1.00	0.25	--	--	--	--	37.75	0.614	0.714	0.329	17.50	0.51	0.83	0.34
	SDRT	218.68	0.60	0.70	0.30	228.97	0.53	0.72	0.25	205.19	0.63	0.73	0.35	63.17	0.75	0.54	0.30
	Tau	185.31	0.59	0.64	0.22	179.14	0.57	0.59	0.16	163.52	0.81	0.51	0.33	63.74	0.51	0.74	0.24
Boys Only	SSRT	--	--	--	--	383.57	0.60	0.82	0.43	--	--	--	--	--	--	--	--
	% Failed Inhibits	36.11	0.21	1.00	0.21	--	--	--	--	34.75	0.83	0.58	0.41	8.50	0.81	0.55	0.36
	SDRT	238.08	0.52	0.89	0.41	185.35	0.77	0.59	0.36	183.00	0.83	0.58	0.41	72.05	0.65	0.74	0.39
	Tau	185.31	0.60	0.72	0.32	157.10	0.65	0.68	0.33	163.52	0.78	0.68	0.47	66.44	0.50	0.86	0.36
Girls Only	SSRT	--	--	--	--	338.05	0.71	0.67	0.38	--	--	--	--	--	--	--	--
	% Failed Inhibits	13.89	0.65	0.73	0.39	--	--	--	--	44.25	0.38	0.94	0.31	17.50	0.57	0.83	0.41
	SDRT	205.94	0.70	0.60	0.30	217.93	0.76	0.55	0.31	216.65	0.67	0.81	0.48	54.61	0.91	0.30	0.21
	Tau	130.01	0.83	0.40	0.23	212.82	0.58	0.64	0.22	206.24	0.50	0.88	0.38	52.70	0.80	0.43	0.23

Table 7. Optimal Cutoff points, and their sensitivity, specificity, and Youden's Index values, for distinguishing individuals with ADHD from typically developing peers.

Overall, sensitivity levels for optimal cutoffs of SSRT, failed inhibit, SDRT, and tau in all samples ranged from .21 - 92, and specificity levels ranged from .30 – 1.0. Though these ranges are wide, they are largely consistent with those reported for well-established behavioral rating scales such as the CBCL Attention Problems subscale (see: Jarrett, Meter, Youngstrom, Hilton, & Ollendick, 2018); this will be addressed further within the discussion.

No single variable appeared to consistently provide the best combination of sensitivity and specificity (i.e. highest Youden's index) across developmental stages though tau consistently provided the lowest Youden's index across samples and tasks. However, there were possible developmental differences, such that within the school-aged FRIENDS and CAL GNG tasks, SDRT was strongest, while in the adults, it was failed inhibits.

Optimal cutoffs selected for SSRT, failed inhibit rate, SDRT, and tau all generally tended to have higher specificity than sensitivity. For example, across samples, the highest Youden's index value for boys and girls together was SDRT in the CAL GNG task (YI = .35). With a sensitivity of .63, specificity of .73, an SDRT cutoff of 217ms would correctly identify 63% of individuals with ADHD as having the disorder, and correctly identify 73% of individuals *without* ADHD as *not having* the disorder. See table 7 for the sensitivity, specificity, and Youden's Index values for other all other cognitive variables within each sample.

Gender differences. In boys, SSRT, SDRT, and tau tended to have the highest Youden's indices. But among girls only % failed inhibits and SDRT yielded the highest Youden's indices.

Post-Hoc Analyses: Including Subclinical Participants

Aim 1: Sensitivity and Specificity. Additional post-hoc analyses were conducted to determine if performance variables continued to significantly discriminate between those with

and without ADHD if individuals with subclinical levels of ADHD symptoms were combined with the typically developing control participants to form a broader “non-ADHD” classification group. Descriptive information for these additional subclinical participants is included in table 8.

	FRIENDS	CAL	B&B
N (Boys:Girls)	28 (16:12)	32 (15:17)	33 (17:16)
Age in years	5.25 (.44)	9.53 (1.16)	20.79 (1.64)
Estimated FSIQ	103.11 (12.84)	111.91 (91)	108.54 (8.69)
Inattention			
Total # of Symptoms	4.14 (2.59)	3.00 (2.38)	1.73 (1.57)
Parent BASC-2	50.82 (6.08)	53.97 (9.78)	--
Teacher BASC-2	--	51.41 (8.83)	--
Parent Conners	49.39 (5.43)	55.31 (8.49)	--
Teacher Conners	63.14 (16.36)	50.84 (9.35)	--
CAARS Inattention	--	--	55.15 (16.12)
Hyperactivity/Impulsivity			
Total # of Symptoms	4.68 (2.31)	2.25 (2.21)	1.58 (1.28)
Parent BASC-2	53.25 (8.67)	52.44 (10.89)	--
Teacher BASC-2	--	49.38 (8.56)	--
Parent Conners	57.00 (7.01)	55.22 (13.52)	--
Teacher Conners	61.04 (11.86)	52.31 (10.41)	--
CAARS Hyperactivity	--	--	50.06 (14.26)
Go-No Go Task			
N (Boys:Girls)	--	16 (8:8)	--
Comission Errors (%)	19.84 (18.17)	33.53 (14.28)	13.45 (8.55)
Reaction Time Variables			
Mean RT	751.00 (148.45)	607.55 (75.41)	356.96 (29.24)
Standard Deviation RT	229.66 (66.04)	187.44 (31.38)	69.90 (12.96)
RT mu	561.27 (192.69)	445.24 (63.78)	294.46 (28.90)
RT sigma	109.57 (82.96)	88.95 (27.15)	30.63 (9.51)
RT tau	189.72 (80.26)	162.31 (33.91)	62.49 (15.744)
Stop Signal Reaction Time Task			
N (Boys:Girls)	--	22 (9:13)	--
SSRT	--	383.72 (127.90)	--
Reaction Time Variables			
Mean RT	--	808.06 (206.10)	--
Standard Deviation RT	--	242.40 (104.01)	--
RT mu	--	607.62 (165.98)	--
RT sigma	--	127.66 (85.13)	--
RT tau	--	200.43 (94.88)	--

**=p<.01; *=p<.05

Table 8. Descriptive information for additional "subclinical" participants, who did not meet full diagnostic criteria but endorsed greater than 3 symptoms of inattention or hyperactivity/ impulsivity. Means, with standard deviation in parentheses. All ratings scales reported in T-scores.

Determining Discriminability. Table 9 provides descriptive statistics for cognitive task variables across all samples. Diagnostic group differences remained largely unchanged when individuals with subclinical symptomatology were combined with typically developing typically developing peers such that failed inhibit rate remained higher and SSRT and SDRT remained longer in those with ADHD, across all tasks.

	FRIENDS			**CAL SSRT**			**CAL GNG**			**B&B GNG**		
	Non-ADHD	***ADHD***	**Test Statistics**	***Non-ADHD***	***ADHD***	**Test Statistics**	***Non-ADHD***	***ADHD***	**Test Statistics**	***Non-ADHD***	***ADHD***	**Test Statistics**
N (Boys:Girls)	61 (34:27)	75 (52:23)	--	98 (43:55)	194 (128:66)	--	51 (27:24)	70 (46:24)	--	105 (59:46)	61 (26:35)	--
SSRT	--	--	--	344.76 (114.20)	441.57 (150.31)	*F*(1,291) = 31.46** η2 = .098	--	--	--	--	--	--
% Failed Inhibits	17.40 (14.50)	23.04 (18.37)	F(1,135) = 3.82^ η2 = .028	--	--	--	32.21 (14.00)	43.49 (15.90)	*F(1,120) = 16.40** η2 = .121*	11.91 (8.57)	21.13 (14.64)	*F*(1, 165) = 26.23** η2 = .138
RT Variables												
mRT	702.06 (142.92)	711.87 (143.92)	F(1,135) = .157 η2 = .001	795.22 (175.67)	821.46 (169.59)	*F(1,291) = 1.52 η2 = .005*	591.11 (82.77)	602.23 (87.37)	*F(1,120) = .50 η2 = .004*	356.20 (35.20)	348.01 (47.21)	*F*(1, 165) = 1.62 η2 = .010
SDRT	213.49 (65.67)	242.93 (71.84)	F(1,135) = 6.10* η2 = .044	216.88 (78.32)	247.95 (80.78)	*F(1,291) = 9.83** η2 = .033*	187.49 (33.92)	219.58 (38.75)	*F(1,120) = 22.45** η2 = .159*	66.20 (12.84)	73.38 (15.82)	*F*(1, 165) = 10.16** η2 = .058
RT mu	520.67 (167.96)	510.11 (170.65)	F(1,135) = .131 η2 = .001	618.65 (158.57)	624.07 (181.56)	*F(1,291) = .063 η2 < .001*	422.98 (68.45)	401.99 (89.07)	*F(1,120) = 1.98 η2 = .016*	298.08 (33.84)	282.52 (41.44)	*F*(1, 165) = 6.97** η2 = .041
RT sigma	95.56 (73.07)	108.98 (77.95)	F(1,135) = 1.06 η2 = .008	115.39 (58.25)	134.12 (72.61)	*F(1,291) = 4.92* η2 = .017*	80.20 (27.75)	93.32 (35.58)	*F(1,120) = 4.80* η2 = .039*	29.87 (10.67)	31.64 (11.65)	*F*(1, 165) = .99 η2 = .006
RT tau	181.38 (77.14)	202.04 (96.63)	F(1,135) = 1.84 η2 = .014	176.57 (96.85)	197.40 (93.78)	*F(1,291) = 3.14 η2 = .011*	168.13 (39.01)	200.24 (47.57)	F(1,120) = 15.59** η2 = .116	58.11 (15.11)	65.49 (16.53)	*F*(1, 165) = 68.57** η2 = .05

Table 9. ADHD vs non-ADHD diagnostic group differences in GNG and SSRT task variables, across all samples. Means, with standard deviation in parentheses.

As would be expected, AUC values dropped slightly across parameters, though results were generally the same (see Table 10). SSRT and failed inhibit rate were still predictive. Of the reaction time distribution variables, SDRT, sigma, and tau also remained discriminatory (See Fig. 9)

		FRIENDS GNG			CAL SSRT			CAL GNG			B&B GNG		
		AUC	Std. Error	Asymp. Sig	AUC	Std. Error	Asymp. Sig	AUC	Std. Error	Asymp. Sig	AUC	Std. Error	Asymp. Sig
Boys and Girls	SSRT	--	--	--	0.694	0.032	<.001**	--	--	--	--	--	--
	% FI	0.583	0.049	0.095	--	--	--	0.702	0.047	<.001**	0.706	0.042	<.001**
	mean RT	0.531	0.050	0.533	0.569	0.036	0.055	0.528	0.053	0.603	0.415	0.047	0.069
	sd RT	0.618	0.048	0.018*	0.635	0.035	<.001**	0.729	0.045	<.001**	0.633	0.045	0.004**
	RT mu	0.480	0.050	0.692	0.506	0.037	0.859	0.385	0.052	0.031*	0.351	0.046	0.001**
	RT sigma	0.545	0.050	0.369	0.586	0.035	0.016*	0.643	0.052	0.007**	0.530	0.047	0.526
	RT tau	0.572	0.049	0.148	0.570	0.035	0.05^	0.711	0.047	<.001**	0.622	0.044	0.009**
Boys Only	SSRT	--	--	--	0.738	0.041	<.001**	--	--	--	--	--	--
	% FI	0.501	0.063	0.993	--	--	--	0.683	0.065	0.009**	0.715	0.057	0.002**
	mean RT	0.612	0.064	0.080	0.555	0.05	0.281	0.543	0.072	0.545	0.456	0.069	0.523
	sd RT	0.628	0.062	0.046*	0.683	0.047	<.001**	0.713	0.061	0.002**	0.686	0.064	0.007**
	RT mu	0.564	0.065	0.318	0.463	0.051	0.468	0.395	0.072	0.137	0.383	0.068	0.088
	RT sigma	0.585	0.065	0.182	0.549	0.052	0.34	0.620	0.070	0.088	0.598	0.070	0.150
	RT tau	0.589	0.062	0.163	0.639	0.049	0.006**	0.732	0.062	<.001**	0.678	0.062	0.009**
Girls Only	SSRT	--	--	--	0.657	0.05	0.003**	--	--	--	--	--	--
	% FI	0.695	0.076	0.019*	--	--	--	0.674	0.078	0.038*	0.701	0.061	0.002**
	mean RT	0.488	0.087	0.884	0.615	0.052	0.03*	0.559	0.085	0.483	0.389	0.065	0.090
	sd RT	0.625	0.082	0.131	0.644	0.05	0.007**	0.769	0.068	0.001**	0.571	0.065	0.277
	RT mu	0.404	0.084	0.247	0.551	0.054	0.333	0.411	0.084	0.293	0.330	0.063	0.009**
	RT sigma	0.492	0.085	0.922	0.615	0.051	0.029*	0.671	0.079	0.042*	0.460	0.065	0.535
	RT tau	0.560	0.084	0.465	0.575	0.052	0.155	0.724	0.073	0.008**	0.553	0.065	0.418

Table 10. Area under the curve statistics for all tasks: FRIENDS GNG, CAL SSRT, CAL GNG, and B&B GNG. Distinguishing between individuals with ADHD and non-ADHD peers (typically developing and subclinical participants combined).

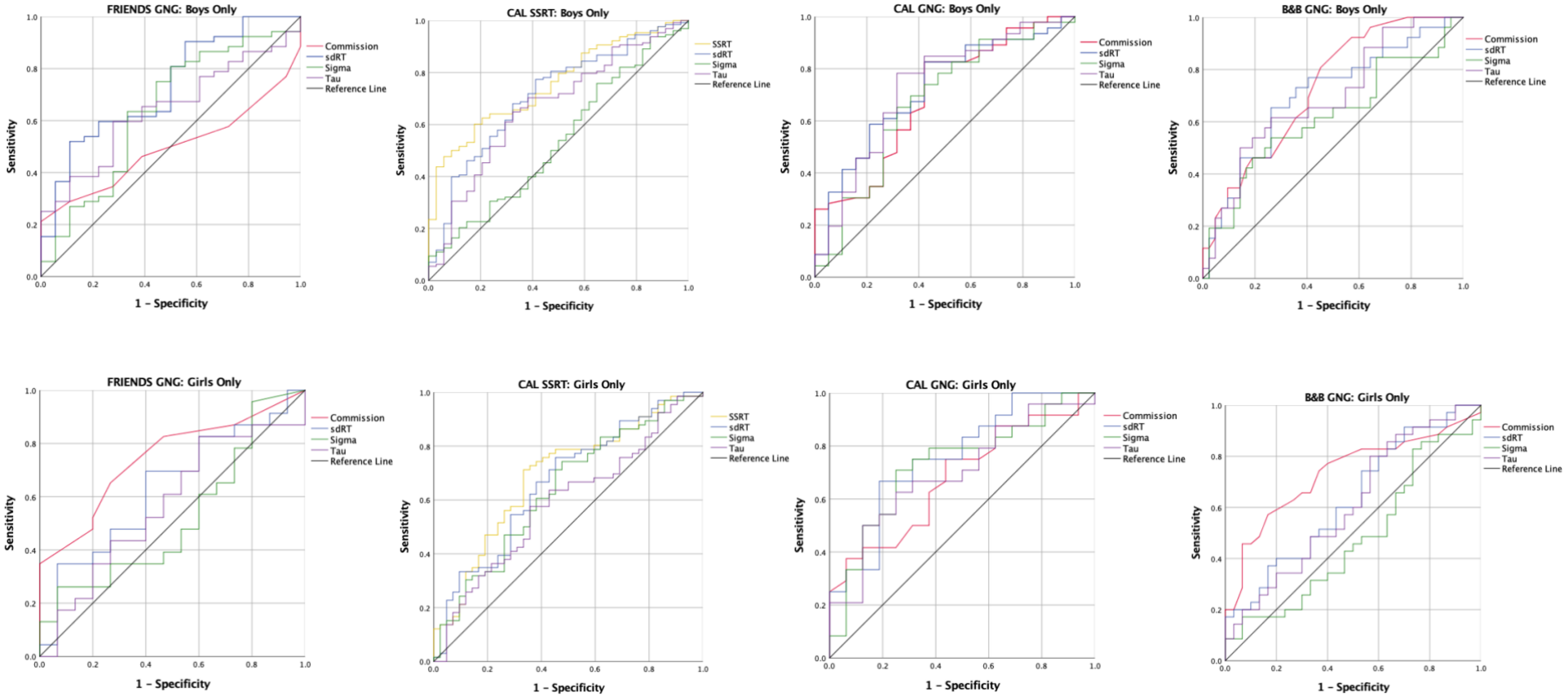

Figure 9. ROC curves for all tasks, distinguishing children with ADHD from typically developing peers, by gender (comparing boys and girls separately).

Gender differences. The pattern of gender differences found when distinguishing between ADHD and typically developing peers also remained generally consistent, such that SSRT and failed inhibit rate remained discriminatory for both boys and girls, SDRT was discriminatory in boys across samples and primarily driven by tau, and in girls SDRT, sigma, and tau were still only discriminatory in the two CAL tasks.

Comparing Predictive Utility. Across all samples, only one comparison was significant: in the CAL SSRT task, SSRT predicted ADHD status better than tau. See Table 11 for details.

		FRIENDS GNG				CAL SSRT				CAL GNG				B&B GNG			
		r Control	r ADHD	z	*p*	r Control	r ADHD	z	*p*	r Control	r ADHD	z	*p*	r Control	r ADHD	z	*p*
Boys and Girls	SSRT vs SDRT	--	--	--	--	0.296	0.337	1.460	0.144	--	--	--	--	--	--	--	--
	SSRT vs tau	--	--	--	--	0.342	0.476	3.330	< 0.001*	--	--	--	--	--	--	--	--
	FI vs SDRT	0.221	0.217	+	+	--	--	--	--	0.16	0.212	-0.45	0.65	0.168	-0.109	1.2	0.23
	FI vs tau	0.427	0.209	+	+	--	--	--	--	0.08	0.127	-0.14	0.89	0.09	-0.092	1.44	0.15
	SDRT vs tau	0.496	0.616	+	+			^		0.913	0.804	0.63	0.53	0.808	0.937	0.45	0.65
Boys Only	SSRT vs SDRT	--	--	--	--	0.374	0.337	1.060	0.289	--	--	--	--	--	--	--	--
	SSRT vs tau	--	--	--	--	0.378	0.530	2.010	0.045*	--	--	--	--	--	--	--	--
	FI vs SDRT	0.242	0.354	+	+	--	--	--	--	0.108	0.208	-0.36	0.72	0.171	-0.022	0.35	0.73
	FI vs tau	0.583	0.232	+	+	--	--	--	--	0.032	0.132	-0.57	0.57	0.03	-0.1	0.45	0.65
	SDRT vs tau	0.364	0.733	+	+	0.742	0.533	1.020	0.306	0.882	0.956	-0.63	0.53	0.694	0.944	0.19	0.85
Girls Only	SSRT vs SDRT	--	--	--	--	0.238	0.373	0.22	0.830	--	--	--	--	--	--	--	--
	SSRT vs tau	--	--	--	--			^		--	--	--	--	--	--	--	--
	FI vs SDRT	0.248	0.002	0.67	0.50	--	--	--	--	0.311	0.369	-1.09	0.28	0.156	-0.15	1.57	0.12
	FI vs tau	0.301	0.195	1.36	0.17	--	--	--	--	0.249	0.168	-0.52	0.61			^	
	SDRT vs tau	0.706	0.396	+	+			^		0.945	0.482	0.77	0.44			^	

^ AUC comparisons were not tested if either variable was not significantly discriminatory.

+ Despite significance, the average of the two AUCs are too low to reliably compare (Hanley & McNeil, 1983).

Table 11. Comparison of AUC values across all tasks. Distinguishing between individuals with ADHD and non-ADHD peers (typically developing and subclinical participants combined).

Gender differences. In boys, across all samples, only one comparison was significant: SSRT predicted ADHD status better than tau. No comparisons were significant in girls in any sample.

Aim 2: Creating Cutoffs.

There were no changes to the pattern of highest Youden's indices across samples or genders when subclinical participants were added; see Table 12.

		FRIENDS GNG				CAL SSRT				CAL GNG				B&B GNG			
		Value	Sens.	Spec.	Youden	Value	Sens.	Spec.	Youden	Value	Sens.	Spec.	Youden	Value	Sens.	Spec.	Youden
Boys and Girls	SSRT	--	--	--	--	445.89	0.47	0.85	0.32	--	--	--	--	--	--	--	--
	% FI	36.11	0.25	0.93	0.19	--	--	--	--	33	0.757	0.569	0.326	16.5	0.557	0.762	0.319
	SDRT	253.29	0.44	0.79	0.23	221.23	0.59	0.63	0.22	205.19	0.63	0.74	0.37	69.54	0.57	0.66	0.23
	Tau	239.80	0.37	0.79	0.16	151.24	0.71	0.42	0.13	163.52	0.81	0.53	0.34	63.74	0.51	0.69	0.19
Boys Only	SSRT	--	--	--	--	383.57	0.60	0.82	0.43	--	--	--	--	--	--	--	--
	% FI	36.11	0.21	0.91	0.12	--	--	--	--	34.75	0.826	0.556	0.382	8.50	0.81	0.54	0.35
	SDRT	149.18	0.90	0.32	0.23	187.83	0.77	0.56	0.32	183.53	0.83	0.56	0.38	72.05	0.65	0.75	0.40
	Tau	185.31	0.60	0.59	0.18	151.24	0.70	0.58	0.29	163.52	0.78	0.67	0.45	61.85	0.62	0.73	0.34
Girls Only	SSRT	--	--	--	--	338.05	0.71	0.67	0.38	--	--	--	--	--	--	--	--
	% FI	36.11	0.35	0.96	0.31	--	--	--	--	44.5	0.375	0.958	0.333	17.5	0.571	0.804	0.376
	SDRT	223.58	0.61	0.67	0.28	220.99	0.74	0.53	0.27	216.65	0.67	0.83	0.50	80.88	0.29	0.87	0.16
	Tau	209.13	0.52	0.70	0.23	212.82	0.58	0.62	0.19	197.93	0.63	0.75	0.38	53.48	0.80	0.35	0.15

Table 12. Optimal Cutoff points, and their sensitivity, specificity, and Youden's Index values, for distinguishing individuals with ADHD from and non-ADHD peers (typically developing and subclinical participants combined).

Discussion

Overall, findings were consistent across samples such that the traditionally used variables SSRT (all AUC > .68) and failed inhibit rate (all AUC > .69), as well as variable reaction time (indexed by SDRT; all AUC > .66) successfully discriminated between individuals with and without ADHD. When RT variability was probed further using ex-Gaussian parameters, it was found that this was generally driven by tau (all AUC > .59) rather than sigma. Though no single variable was found to consistently provide the greatest combination of sensitivity and specificity when using Youden's Index, optimal cutoffs generally showed higher specificity than sensitivity.

Comparative Strength of Discriminability

AUC values for behavioral rating scales are typically much higher than those reported here for cognitive performance. For example, in samples of children across multiple nationalities and derived from both community and clinic-referred populations, AUCs for the CBCL Attention Problems Subscale is reported to be between .74 and .96 (Chen, Faraone, Biederman, & Tsuang, 1994; de la Osa, Granero, Trepat, Domenech, & Ezpeleta, 2016; Hudziak, Copeland, Stanger, & Wadsworth, 2004; Lampert, Polanczyk, Tramontina, Mardini, & Rohde, 2004; Raiker et al., 2017; Rey, Morris-Yates, & Stanislaw, 1992). Similarly, for adults, the CAARS, ADHD-RS, and ASRS have also shown high AUC values, between .75 - .9 (Kessler et al., 2007; Sadeghi-Bazargani et al., 2014; Szomlaiski et al., 2009). However, this advantage is largely tautological as rating scales are specifically designed to reflect signs and symptoms of any the disorder(s) they address.

Performance on cognitive tasks has frequently shown diagnostic discriminability above chance. However, due to the lack of the tautological advantage, with a few exceptions (see: Berger et al., 2017; Leth-Steensen et al., 2000) their AUC values are generally lower than those

for questionnaires (Devena & Watkins, 2012; Doyle et al., 2000; Jarrett et al., 2018; Pineda, Puerta, Aguirre, García-Barrera, & Kamphaus, 2007; Teicher et al., 2012). That being said, the strength of using cognitive tasks lies in their ability to speak to the possible causal and transdiagnostic mechanisms that may contribute to the development or maintenance of disorder. For example, studies of the Conners CPT-II, which is similar to the GNG tasks included within the current study, found AUC values for standard error and failed inhibit between .63 - .71 (Jarrett et al., 2018; Teicher et al., 2012). Other studies, which examined traditional paper-and-pencil neuropsychological tasks (but did not include reaction time tasks), found AUC values equivalent or somewhat lower than the current study, varying from .56 - .7 (Devena & Watkins, 2012; Doyle et al., 2000; Pineda et al., 2007). Thus, the AUC values for SSRT, failed inhibit rate, SDRT, and tau found within the current study are comparable to those found in tasks which are currently used widely throughout research and clinical practice and point to specific cognitive dysfunctions.

Continuity Across Developmental Periods

The current study included three specific developmental periods (i.e., preschool/ kindergarten-age, middle childhood, and early adulthood), as well as both male and female participants, which allowed for the evaluation of possible developmental- and gender-based differences. When measured using standard deviation, variability in reaction time discriminated between individuals with and without ADHD across all age-groups. When variability was further explored using the ex-Gaussian parameters, the differences in SDRT were driven primarily by tau in both the school-aged and adult samples, but not the preschool sample. That is, the shape of speeded reaction time performance among preschool-aged children with ADHD was more consistently variable across the distribution of performance, and they did not demonstrate

the characteristic longer tail that has been commonly documented in school-aged children and adults with ADHD. This could indicate that though ex-Gaussian parameters are more accurate descriptors of the RT distribution in older individuals, assessments of preschool/kindergarten-age children should continue to use SDRT as the primary method of measuring impaired speed.

However, it's likely that the lack of discriminability found in sigma/tau within this sample instead reflects sample-specific methodological differences. To date, the only other known study examining differences in ex-Gaussian parameters of reaction time within preschoolers at risk for ADHD found that sigma and tau were greater in children with ADHD than typically developing controls when using an ISI of 3000 milliseconds but not with an ISI of 1500 milliseconds (Hwang-Gu et al., 2019). The authors of that study posited that impairments in processing speed may only be present in preschoolers with ADHD when engaged in slower-paced tasks which elicit less cognitive activation. The current study's task included an ISI of 1000 milliseconds, which, while standard among school aged and adolescents/adults, may nevertheless have been rapid enough to elicit greater cognitive activation and prevent lapses of performance among preschool aged children.

Additionally, the sigma and tau parameters generated using empirical data produced a simulated RT distribution that was significantly correlated with the empirical distribution (r=.4), but was still well below the commonly applied threshold (r = .8), suggesting that the parameters that were output may not accurately reflect the data. This is likely due to the smaller number of trials in the GNG task employed within the FRIENDS sample (i.e. 60). Because data can be best fit to a distributional model when using as many individual data points as possible (Lacouture & Cousineau, 2008), the poor model fit within this task may be due to the limited number of trials used to derived the model parameters, rather than because RT distributions in preschool/

kindergarten-aged children don't generally fit the ex-Gaussian shape. Hwang-Gu et al. (2019)'s GNG task had 200 total trials, which would be expected to yield better fits, though fit was not reported. Therefore, the lack of discriminability for tau found within the FRIENDS sample is likely a methodological limitation of this task rather than an indication that RT variability functions differently in young children.

Differing Task Performance by Gender

Two gender differences within the current results also warrant further discussion. First, consistent with previous findings from a similar study of school-aged boys (Leth-Steensen et al., 2000), SDRT and tau were both found to discriminate between ADHD and non-ADHD controls within the two CAL tasks, regardless of gender. However, within the B&B sample, SDRT and tau were only discriminatory in young men, not young women. The current study therefore builds upon previous work by replicating the results found in male participants, while also suggesting that cognitive performance may operate differently in young women. Though this gender difference was not expected within the current study, two possible explanations warrant consideration.

First, normative studies of adult cognition have found greater variability of reaction time in healthy young women than healthy young men (Deary & Der, 2005; Dykiert, Der, Starr, & Deary, 2012; Reimers & Maylor, 2006); if healthy young women have more variable performance and speed within RT tasks, then standard deviation or tau may not be useful markers to distinguish between young women with and without ADHD. Second, differences in symptom presentation (e.g., fewer hyperactive/impulsive symptoms) and comorbidity (higher rates of depression and anxiety) found in women with ADHD (Gershon, 2002; Rasmussen & Levander, 2009; Wilens et al., 2009) may also impact cognitive performance and result in

different performance profiles than their male peers. Therefore, as much as replication with additional and larger samples is essential to the advancement of scientific findings, the unique impairment profiles found within the men and women in the B&B sample highlight the need for research which is inclusive of all demographic groups rather than assuming that patterns found in male samples will replicate in female samples. This is specifically important within ADHD research, because though the field continues to be dominated by studies about boys (in large part because they are more likely to present in clinics), research finds that girls with ADHD also face a range of impairments (Gershon, 2002; Hinshaw, 2002; Lee & Hinshaw, 2006).

Additional gender differences were found in the two child samples (FRIENDS and CAL). Though failed inhibit rate was consistently discriminatory in both men and women in the B&B sample, it was discriminatory in boys but not girls in the CAL sample, and for girls but not boys in the FRIENDS sample. This may suggest that the mechanisms driving broader cognitive deficits in ADHD may differ by gender or by developmental stage, and that any normative values or cutoff scores used clinically should be developed for each gender separately. Though it's also possible these differences may be a reflection of limited power in the current samples.

Optimal Cutoffs.

Across all four tasks and with few exceptions, optimal cutoffs for SSRT, failed inhibit rate, SDRT, and tau had greater sensitivity and specificity than chance, and were equal to or stronger than values found in studies of similar tasks frequently used in current clinical practice. Of particular importance was the strength of SSRT, which is not frequently used within clinical evaluations, despite being frequently employed within research regarding cognitive performance within ADHD (Crosbie et al., 2013; Nigg et al., 2018). Therefore, the current study's results indicate that SSRT, SDRT, and tau may perform as well as commission rate, and that adding

SSRT as an additional measure of inhibition and SDRT/tau as indicators of performance speed would likely contribute to the breadth of information gained within an evaluation. A recent study using ROC analyses to evaluate the ability of cognitive performance to predict ADHD diagnosis found very strong discriminative ability (CPT Total Index AUC = .91 - .96), with the Youden's Index ranging from .81 - .91 (Berger et al., 2017). Though our values were somewhat weaker than those previously reported, this would be expected as they used a clinically referred sample rather than community recruitment.

However, the definition of "optimal" varies greatly based on a) the way a tool is intended to be used, and b) the possible consequences of incorrect discrimination (Berger et al., 2017). In regards to the former, this would require determining if a measure is more suited to function as an early screener or later rule-out. Screening tools generally select a relatively low optimal cutoff based on setting a high sensitivity level to maximize the number of individuals warranting further evaluation, while tools used for ruling out a diagnosis use higher cutoffs based on setting a high specificity level. In regards to the latter, possible consequences of a "false positive" diagnosis of ADHD could include social stigma associated with diagnosis, failure to make an appropriate alternative diagnosis, or even possible iatrogenic effects of an unnecessary medication (Abramovitch, 2016; Birnbaum et al., 2005; Evans et al., 2010; Wiener et al., 2012), while possible consequences of a "false negative" missed diagnosis could include severe outcomes associated with severe or untreated ADHD such as school failure, dangerous or sensation-seeking behavior, illegal drug use, and increased peer and familial conflict (Biederman et al., 2004b; Birnbaum et al., 2005; Eakin et al., 2004; D. W. Goodman, 2007; Harpin, 2005; Minde et al., 2003). Given these criteria, cognitive tasks such as those within the current study are likely not suited to functioning as screeners as they cannot be quickly and cheaply

administered like behavioral questionnaires. Additionally, because a large portion of people meeting diagnostic criteria do not show impairments on cognitive tasks such as these, their use as a method to rule out a diagnosis would likely result in many missed diagnoses. Therefore, and consistent with the arguments made by several previous studies evaluating similar tasks (Doyle et al., 2000; Faraone et al., 2016), cutoffs may be most valuable when used to explore tasks' strengths and weaknesses as a descriptive tool for functional evaluations, rather than as a diagnostic tool.

Limitations and Future Directions

It is important to note that while the same broad GNG paradigm was implemented with all three age-groups, the specific tasks administered did vary in several ways, including type and number of stimuli, complexity, time between stimuli, and feedback regarding accuracy. Though each task was designed for the age-range with which it was used, it's possible that slight variations in the demands of each task might have led to some of the variation in performance, rather than age-related differences in underlying cognitive mechanisms. Future multi-generational cross-sectional studies using more standardized age-specific GNG tasks, or longitudinal studies through childhood to early adulthood may be better able to answer this question. Furthermore, gender differences in discriminability may be due to reduced power within each subsample, particularly given the relatively few number of girls/women, rather than genuine differences in cognition across genders. Therefore, future studies should aim for a more even balance of male and female participants, enabling more careful evaluation of if and how they demonstrate genuinely different cognitive performance profiles, requiring separate normative values.

Summary and Conclusions

The current study had two primary goals. First, to determine if ex-Gaussian parameters of reaction time are able discriminate between children and young adults with and without ADHD as well as more traditional methods of indexing cognition. Second, to examine the sensitivity and specificity of these cognitive measures to elucidate how they might best contribute to mechanistic research as well as clinical practice.

Overall, SSRT, failed inhibit rate, and SDRT predicted ADHD diagnosis better than chance. When RT distributions were probed further using ex-Gaussian parameters, the effect of variability was found to be driven primarily by tau, though this varied somewhat by age group and was more consistent in boys than in girls. When optimal cutoff values were determined using Youden's Index, no single variable appeared to consistently provide the best combination of sensitivity and specificity, though sensitivity and specificity were generally all above chance. These findings suggest measures of reaction time can discriminate individuals with ADHD as well as more traditional measures such as SSRT and failed inhibit rate, and particularly, that using ex-Gaussian parameters provides a better description of variability in performance during speeded tasks than mean/standard deviation. Therefore, assessment tools used both in research and clinical practice would benefit from implementing these methods.

References

Abney, D. H., McBride, D. M., & Petrella, S. N. (2013). Interactive effects in transfer-appropriate processing for event-based prospective memory: The roles of effort, ongoing task, and PM cue properties. *Memory & Cognition, 41*(7), 1032-1045. doi:10.3758/s13421-013-0324-7

Abramovitch, A. (2016). Misdiagnosis of ADHD in individuals diagnosed with obsessive-compulsive disorder: Guidelines for practitioners. *Current Treatment Options in Psychiatry, 3*(3), 225-234.

Achenbach, T. M., & Rescorla, L. A. (2003). *Manual for the ASEBA Adult Forms & Profiles*. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.

Adamo, N., Huo, L., Adelsberg, S., Petkova, E., Castellanos, F. X., & Di Martino, A. (2014). Response time intra-subject variability: commonalities between children with autism spectrum disorders and children with ADHD. *European Child & Adolescent Psychiatry, 23*(2), 69-79. doi:10.1007/s00787-013-0428-4

Alderson, R. M., Rapport, M. D., & Kofler, M. J. (2007). Attention-Deficit/Hyperactivity disorder and behavioral inhibition: A meta-analytic review of the stop-signal paradigm. *J Abnorm Child Psychol, 35*(5), 745-758. doi:10.1007/s10802-007-9131-6

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

Antonini, T. N., Narad, M. E., Langberg, J. M., & Epstein, J. N. (2013). Behavioral correlates of reaction time variability in children with and without ADHD. *Neuropsychology, 27*(2), 201-209. doi:10.1037/a0032071

Balota, D. A., & Spieler, D. H. (1999). Word frequency, repetition, and lexicality effects in word recognition tasks: Beyond measures of central tendency. *Journal of Experimental Psychology-General, 128*(1), 32-55. doi:10.1037//0096-3445.128.1.32

Barkley, R. A. (1997). Behavioral inhibition, sustained attention, and executive functions: Constructing a unifying theory of ADHD. *Psychological Bulletin, 121*(1), 65-94.

Barrouillet, P., Gavens, N., Vergauwe, E., Gaillard, V., & Camos, V. (2009). Working memory span development: a time-based resource-sharing model account. *Dev Psychol, 45*(2), 477-490. doi:10.1037/a0014615

Bayliss, D. M., Jarrold, C., Baddeley, A. D., Gunn, D. M., & Leigh, E. (2005). Mapping the developmental constraints on working memory span performance. *Dev Psychol, 41*(4), 579-597. doi:10.1037/0012-1649.41.4.579

Bellgrove, M. A., Hawi, Z., Gill, M., & Robertson, I. H. (2006). The cognitive genetics of attention deficit hyperactivity disorder (ADHD): Sustained attention as a candidate phenotype. *Cortex, 42*(6), 838-845. doi:10.1016/s0010-9452(08)70426-x

Benca, C. E., Derringer, J. L., Corley, R. P., Young, S. E., Keller, M. C., Hewitt, J. K., & Friedman, N. P. (2017). Predicting Cognitive Executive Functioning with Polygenic Risk Scores for Psychiatric Disorders. *Behavior Genetics, 47*(1), 11-24. doi:10.1007/s10519-016-9814-2

Berger, I., Slobodin, O., & Cassuto, H. (2017). Usefulness and validity of continuous performance tests in the diagnosis of attention-deficit hyperactivity disorder children. *Archives of Clinical Neuropsychology, 32*(1), 81-93.

Biederman, J., Faraone, S. V., Keenan, K., Benjamin, J., Krifcher, B., Moore, C., . . . Tsuang, M. T. (1992). Further evidence for family-genetic risk-factors in Attention-Deficit Hyperactivity Disorder – patterns of comorbidity in probands and relatives in psychiatrically and pediatrically referred samples. *Archives of General Psychiatry, 49*(9), 728-738.

Biederman, J., Monuteaux, M. C., Doyle, A. E., Seidman, L. J., Wilens, T. E., Ferrero, F., . . . Faraone, S. V. (2004a). Impact of executive function deficits and attention-deficit/hyperactivity disorder (ADHD) on academic outcomes in children. *J Consult Clin Psychol, 72*(5), 757-766. doi:10.1037/0022-006X.72.5.757

Biederman, J., Monuteaux, M. C., Doyle, A. E., Seidman, L. J., Wilens, T. E., Ferrero, F., . . . Faraone, S. V. (2004b). Impact of executive function deficits and Attention-Deficit/Hyperactivity Disorder (ADHD) on academic outcomes in children. *Journal of Consulting and Clinical Psychology, 72*(5), 757-766. doi:10.1037/0022-006X.72.5.757

Birnbaum, H. G., Kessler, R. C., Lowe, S. W., Secnik, K., Greenberg, P. E., Leong, S. A., & Swensen, A. R. (2005). Costs of attention deficit-hyperactivity disorder (ADHD) in the US: excess costs of persons with ADHD and their family members in 2000. *Current Medical Research and Opinion, 21*(2), 195-205. doi:10.1185/030079904x20303

Bitsakou, P., Psychogiou, L., Thompson, M., & Sonuga-Barke, E. J. S. (2009). Delay Aversion in Attention Deficit/Hyperactivity Disorder: An empirical investigation of the broader phenotype. *Neuropsychologia, 47*(2), 446-456. doi:10.1016/j.neuropsychologia.2008.09.015

Boonstra, A. M., Kooij, J. J. S., Oosterlaan, J., Sergeant, J. A., & Buitelaar, J. K. (2010). To Act or Not to Act, That's the Problem: Primarily Inhibition Difficulties in Adult ADHD. *Neuropsychology, 24*(2), 209-221. doi:10.1037/a0017670

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.

Borella, E., de Ribaupierre, A., Cornoldi, C., & Chicherio, C. (2013). Beyond interference control impairment in ADHD: evidence from increased intraindividual variability in the color-stroop test. *Child Neuropsychology, 19*(5), 495-515. doi:10.1080/09297049.2012.696603

Braun, J. M., Kahn, R. S., Froehlich, T., Auinger, P., & Lanphear, B. P. (2006). Exposures to environmental toxicants and attention deficit hyperactivity disorder in US children. *Environmental Health Perspectives, 114*(12), 1904-1909. doi:10.1289/ehp.9478

Buzy, W. M., Medoff, D. R., & Schweitzer, J. B. (2009). Intra-individual variability among children with ADHD on a working memory task: an ex-Gaussian approach. *Child Neuropsychology, 15*(5), 441-459. doi:10.1080/09297040802646991

Castellanos, F. X., Sonuga-Barke, E. J., Scheres, A., Di Martino, A., Hyde, C., & Walters, J. R. (2005). Varieties of attention-deficit/hyperactivity disorder-related intra-individual variability. *Biological Psychiatry, 57*(11), 1416-1423. doi:10.1016/j.biopsych.2004.12.005

Castellanos, F. X., & Tannock, R. (2002). Neuroscience of attention-deficit/hyperactivity disorder: The search for endophenotypes. *Nature Reviews Neuroscience, 3*(8), 617-628. doi:10.1038/nrn896

Cepeda, N. J., Blackwell, K. A., & Munakata, Y. (2013). Speed isn't everything: Complex processing speed measures mask individual differences and developmental changes in executive control. *Dev Sci, 16*(2), 269-286. doi:10.1111/desc.12024

Chen, W. J., Faraone, S. V., Biederman, J., & Tsuang, M. T. (1994). Diagnostic accuracy of the Child Behavior Checklist scales for attention-deficit hyperactivity disorder: a receiver-operating characteristic analysis. *Journal of Consulting and Clinical Psychology, 62*(5), 1017.

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

Conners, C. K. (2004). *Conners Continuous Performance Test II*. Toronto, ON: MHS.

Conners, C. K. (2008). *Conners' Rating Scales—3 Technical Manual*. NY: Multi-Health Systems Inc.

Conners, C. K., Epstein, J., & Johnson, D. E. (2001). *Conner's Adult ADHD Diagonostic Interview for DSM-IV: CAADID*: MHS.

Crosbie, J., Arnold, P., Paterson, A., Swanson, J., Dupuis, A., Li, X., . . . Schachar, R. J. (2013). Response Inhibition and ADHD Traits: Correlates and Heritability in a Community Sample. *J Abnorm Child Psychol, 41*(3), 497-507. doi:10.1007/s10802-012-9693-9

Crosbie, J., Perusse, D., Barr, C. L., & Schachar, R. J. (2008). Validating psychiatric endophenotypes: Inhibitory control and attention deficit hyperactivity disorder. *Neuroscience and Biobehavioral Reviews, 32*(1), 40-55. doi:10.1016/j.neubiorev.2007.05.002

Crosbie, J., & Schachar, R. (2001). Deficient inhibition as a marker for familial ADHD. *American Journal of Psychiatry, 158*(11), 1884-1890. doi:10.1176/appi.ajp.158.11.1884

Curchack-Lichtin, J. T., Chacko, A., & Halperin, J. M. (2014). Changes in ADHD Symptom Endorsement: Preschool to School Age. *J Abnorm Child Psychol, 42*(6), 993-1004. doi:10.1007/s10802-013-9834-9

Dawson, M. R. W. (1988). Fitting the ex-Gaussian equation to reaction time distributions. *Behavior Research Methods Instruments & Computers, 20*(1), 54-57. doi:10.3758/bf03202603

de la Osa, N., Granero, R., Trepat, E., Domenech, J. M., & Ezpeleta, L. (2016). The discriminative capacity of CBCL/1½-5-DSM5 scales to identify disruptive and internalizing disorders in preschool children. *European Child & Adolescent Psychiatry, 25*(1), 17-23.

Deary, I. J., & Der, G. (2005). Reaction time, age, and cognitive ability: Longitudinal findings from age 16 to 63 years in representative population samples. *Aging Neuropsychology and Cognition, 12*(2), 187-215. doi:10.1080/13825805990969235

Devena, S. E., & Watkins, M. W. (2012). Diagnostic utility of WISC-IV general abilities index and cognitive proficiency index difference scores among children with ADHD. *Journal of Applied School Psychology, 28*(2), 133-154.

Doyle, A. E., Biederman, J., Seidman, L. J., Weber, W., & Faraone, S. V. (2000). Diagnostic efficiency of neuropsychological test scores for discriminating boys with and without attention deficit-hyperactivity disorder. *Journal of Consulting and Clinical Psychology, 68*(3), 477-488. doi:10.1037//0022-006x.68.3.477

Doyle, A. E., Faraone, S. V., Seidman, L. J., Willcutt, E. G., Nigg, J. T., Waldman, I. D., . . . Biederman, J. (2005). Are endophenotypes based on measures of executive functions useful for molecular genetic studies of ADHD? *Journal of child Psychology and Psychiatry, 46*(7), 774-803. doi:10.1111/j.1469-7610.2005.01476.x

DuPaul, G. J., Anastopoulos, A. D., Power, T. J., Reid, R., Ikeda, M. J., & McGoey, K. E. (1998). Parent ratings of attention-deficit/hyperactivity disorder symptoms: Factor

structure and normative data. *Journal of Psychopathology and Behavioral Assessment, 20*(1), 83-102.

DuPaul, G. J., McGoey, K. E., Eckert, T. L., & VanBrakle, J. (2001). Preschool children with Attention-Deficit/Hyperactivity Disorder: Impairments in behavioral, social, and school functioning. *Journal of the American Academy of Child Adolescent Psychiatry, 40*(5), 515.

Dykiert, D., Der, G., Starr, J. M., & Deary, I. J. (2012). Sex differences in reaction time mean and intraindividual variability across the life span. *Dev Psychol, 48*(5), 1262.

Eakin, L., Minde, K., Hechtman, L., Ochs, E., Krane, E., Bouffard, R., . . . Looper, K. (2004). The marital and family functioning of adults with ADHD and their spouses. *Journal of Attention Disorders, 8*(1), 1-10.

Edwards, M. C., Gardner, E. S., Chelonis, J. J., Schulz, E. G., Flake, R. A., & Diaz, P. F. (2007). Estimates of the validity and utility of the Conners' continuous performance test in the assessment of inattentive and/or hyperactive-impulsive behaviors in children. *J Abnorm Child Psychol, 35*(3), 393-404. doi:10.1007/s10802-007-9098-3

Epstein, J. N., Langberg, J. M., Rosen, P. J., Graham, A., Narad, M. E., Antonini, T. N., . . . Altaye, M. (2011a). Evidence for higher reaction time variability for children with ADHD on a range of cognitive tasks including reward and event rate manipulations. *Neuropsychology, 25*(4), 427-441. doi:10.1037/a0022155

Epstein, J. N., Langberg, J. M., Rosen, P. J., Graham, A., Narad, M. E., Antonini, T. N., . . . Altaye, M. (2011b). Supplemental Material for Evidence for Higher Reaction Time Variability for Children With ADHD on a Range of Cognitive Tasks Including Reward and Event Rate Manipulations. *Neuropsychology, 25*(4), 427-441. doi:10.1037/a0022155.supp

Evans, W. N., Morrill, M. S., & Parente, S. T. (2010). Measuring inappropriate medical diagnosis and treatment in survey data: The case of ADHD among school-age children. *Journal of health economics, 29*(5), 657-673.

Fair, D. A., Bathula, D., Nikolas, M. A., & Nigg, J. T. (2012). Distinct neuropsychological subgroups in typically developing youth inform heterogeneity in children with ADHD. *Proceedings of the National Academy of Sciences of the United States of America, 109*(17), 6769-6774. doi:10.1073/pnas.1115365109

Fallgatter, A. J., Ehlis, A. C., & Herrmann, M. J. (2004). Electrophysiological dysfunction in the anterior cingulate cortex as an endophenotype for attention deficit hyperactivity disorder (ADHD). *European Psychiatry, 19*, 103S-104S.

Faraone, S. V., Newcorn, J. H., Antshel, K. M., Adler, L., Roots, K., & Heller, M. (2016). The Groundskeeper Gaming Platform as a diagnostic tool for Attention-Deficit/Hyperactivity Disorder: Sensitivity, specificity, and relation to other measures. *Journal of Child and Adolescent Psychopharmacology, 26*(8), 672-685. doi:10.1089/cap.2015.0174

Ferguson, A. N., & Bowey, J. A. (2005). Global processing speed as a mediator of developmental changes in children's auditory memory span. *J Exp Child Psychol, 91*(2), 89-112. doi:10.1016/j.jecp.2004.12.006

Florkowski, C. M. (2008). Sensitivity, specificity, receiver-operating characteristic (ROC) curves and likelihood ratios: communicating the performance of diagnostic tests. *The Clinical Biochemist Reviews, 29*(Suppl 1), S83.

Fried, R., Petty, C., Faraone, S. V., Hyder, L. L., Day, H., & Biederman, J. (2016). Is ADHD a risk factor for high school dropout? A controlled study. *Journal of Attention Disorders, 20*(5), 383-389.

Froehlich, T. E., Lanphear, B. P., Auinger, P., Hornung, R., Epstein, J. N., Braun, J., & Kahn, R. S. (2009). Association of tobacco and lead exposures with Attention-Deficit/Hyperactivity Disorder. *Pediatrics, 124*(6), E1054-E1063. doi:10.1542/peds.2009-0738

Fry, A. F., & Hale, S. (2000). Relationships among processing speed, working memory, and fluid intelligence in children. *Biological Psychology, 54*, 34

Gaillard, V., Barrouillet, P., Jarrold, C., & Camos, V. (2011). Developmental differences in working memory: Where do they come from? *J Exp Child Psychol, 110*(3), 469-479. doi:10.1016/j.jecp.2011.05.004

Gallo, E. F., & Posner, J. (2016). Moving towards causality in attention-deficit hyperactivity disorder: overview of neural and genetic mechanisms. *Lancet Psychiatry, 3*(6), 555-567. doi:10.1016/s2215-0366(16)00096-1

Galloway-Long, H., & Huang-Pollock, C. L. (2018). Using inspection time and ex-Gaussian parameters of reaction time to predict executive functions in children with ADHD. *Intelligence, 69*, 186-194.

Gershon, J. (2002). A meta-analytic review of gender differences in ADHD. *Journal of Attention Disorders, 5*(3), 143-154.

Gmehlin, D., Fuermaier, A. B. M., Walther, S., Debelak, R., Rentrop, M., Westermann, C., . . . Aschenbrenner, S. (2014). Intraindividual Variability in Inhibitory Function in Adults with ADHD - An Ex-Gaussian Approach. *PLoS ONE, 9*(12), 19. doi:10.1371/journal.pone.0112298

Golden, C. J., Moses Jr, J. A., Fishburne, F. J., Engum, E., Lewis, G. P., Wisniewski, A. M., . . . Graber, B. (1981). Cross-validation of the Luria-Nebraska Neuropsychological Battery for the presence, lateralization, and localization of brain damage. *Journal of Consulting and Clinical Psychology, 49*(4), 491.

Gomez, P., Ratcliff, R., & Perea, M. (2007). A model of the go/no-go task. *Journal of Experimental Psychology: General, 136*(3), 389.

Goodman, D. W. (2007). The consequences of attention-deficit/hyperactivity disorder in adults. *Journal of Psychiatric Practice®, 13*(5), 318-327.

Goodman, R., & Scott, S. (1999). Comparing the Strengths and Difficulties Questionnaire and the Child Behavior Checklist: is small beautiful? *J Abnorm Child Psychol, 27*(1), 17-24.

Gordon, B., & Carson, K. (1990). The basis for choice reaction-time slowing in Alzheimers Disease. *Brain and Cognition, 13*(2), 148-166. doi:10.1016/0278-2626(90)90047-r

Gottesman, II, & Gould, T. D. (2003). The endophenotype concept in psychiatry: Etymology and strategic intentions. *American Journal of Psychiatry, 160*(4), 636-645. doi:10.1176/appi.ajp.160.4.636

Gu, S. L. H., Gau, S. S. F., Tzang, S. W., & Hsu, W. Y. (2013). The ex-Gaussian distribution of reaction times in adolescents with attention-deficit/hyperactivity disorder. *Res Dev Disabil, 34*(11), 3709-3719. doi:10.1016/j.ridd.2013.07.025

Gudjonsson, G. H., Sigurdsson, J. F., Sigfusdottir, I. D., & Young, S. (2012). An epidemiological study of ADHD symptoms among young persons and the relationship with cigarette smoking, alcohol consumption and illicit drug use. *Journal of child Psychology and Psychiatry, 53*(3), 304-312.

Hajian-Tilaki, K. (2013). Receiver Operating Characteristic (ROC) curve analysis for medical diagnostic test evaluation. *Caspian Journal of Internal Medicine, 4*(2), 627-635.

Hajian-Tilaki, K. (2018). The choice of methods in determining the optimal cut-off value for quantitative diagnostic test evaluation. *Stat Methods Med Res, 27*(8), 2374-2383. doi:10.1177/0962280216680383

Hajian-Tilaki, K., Hanley, J. A., & Nassiri, V. (2011). An extension of parametric ROC analysis for calculating diagnostic accuracy when underlying distributions are mixture of Gaussian. *Journal of Applied Statistics, 38*(9), 2009-2022. doi:10.1080/02664763.2010.545109

Hanley, J. A. (1989). Receiver operating characteristic (ROC) methodology – the state of the art. *Critical Reviews in Diagnostic Imaging, 29*(3), 307-335.

Hanley, J. A., & McNeil, B. J. (1983). A METHOD OF COMPARING THE AREAS UNDER RECEIVER OPERATING CHARACTERISTIC CURVES DERIVED FROM THE SAME CASES. *Radiology, 148*(3), 839-843. doi:10.1148/radiology.148.3.6878708

Harpin, V. A. (2005). The effect of ADHD on the life of an individual, their family, and community from preschool to adult life. *Archives of Disease in Childhood, 90*(suppl 1), i2-i7.

Henry, S. K., Grant, M. M., & Cropsey, K. L. (2018). Determining the optimal clinical cutoff on the CES-D for depression in a community corrections sample. *Journal of Affective Disorders, 234*, 270-275. doi:10.1016/j.jad.2018.02.071

Hervey, A. S., Epstein, J. N., Curry, J. F., Tonev, S., Eugene Arnold, L., Conners, C. K., . . . Hechtman, L. (2006). Reaction time distribution analysis of neuropsychological performance in an ADHD sample. *Child Neuropsychology, 12*(2), 125-140. doi:10.1080/09297040500499081

Hervey, A. S., Epstein, J. N., Curry, J. F., Tonev, S., Eugene Arnold, L., Keith Conners, C., . . . Hechtman, L. (2006). Reaction time distribution analysis of neuropsychological performance in an ADHD sample. *Child Neuropsychol, 12*(2), 125-140. doi:10.1080/09297040500499081

Hinshaw, S. P. (2002). Preadolescent girls with attention-deficit/hyperactivity disorder: I. Background characteristics, comorbidity, cognitive and social functioning, and parenting practices. *Journal of Consulting and Clinical Psychology, 70*(5), 1086.

Hockley, W. E. (1984). Analysis of response-time distributions in the study of cognitive processes. *Journal of Experimental Psychology-Learning Memory and Cognition, 10*(4), 598-615. doi:10.1037/0278-7393.10.4.598

Homack, S., & Riccio, C. A. (2004). A meta-analysis of the sensitivity and specificity of the Stroop Color and Word Test with children. *Archives of Clinical Neuropsychology, 19*(6), 725-743.

Huang-Pollock, C. L., Karalunas, S. L., Tam, H., & Moore, A. N. (2012). Evaluating Vigilance Deficits in ADHD: A Meta-Analysis of CPT Performance. *J Abnorm Psychol, 121*(2), 360-371. doi:10.1037/a0027205

Huang-Pollock, C. L., & Nigg, J. T. (2003). Searching for the attention deficit in attention deficit hyperactivity disorder: The case of visuospatial orienting. *Clinical Psychology Review, 23*(6), 801-830.

Hudziak, J. J., Copeland, W., Stanger, C., & Wadsworth, M. (2004). Screening for DSM-IV externalizing disorders with the Child Behavior Checklist: a receiver-operating

characteristic analysis. *Journal of child Psychology and Psychiatry, 45*(7), 1299-1307. doi:10.1111/j.1469-7610.2004.00314.x

Hutchinson, M., Bradbury, J., Browne, G., & Hurley, J. (2017). Determining the optimal cut-off scores for the Workplace Bullying Inventory. *Nurse Researcher, 25*(3), 46-50. doi:10.7748/nr.2017.e1543

Hwang-Gu, S.-L., Chen, Y.-C., Liang, S. H.-Y., Ni, H.-C., Lin, H.-Y., Lin, C.-F., & Gau, S. S.-F. (2019). Exploring the Variability in Reaction Times of Preschoolers at Risk of Attention-Deficit/Hyperactivity Disorder: an ex-Gaussian Analysis. *J Abnorm Child Psychol*, 1-12.

Jacobson, L. A., Ryan, M., Martin, R. B., Ewen, J., Mostofsky, S. H., Denckla, M. B., & Mahone, E. M. (2011). Working memory influences processing speed and reading fluency in ADHD. *Child Neuropsychology, 17*(3), 209-224. doi:10.1080/09297049.2010.532204

Jarrett, M. A., Meter, A. V., Youngstrom, E. A., Hilton, D. C., & Ollendick, T. H. (2018). Evidence-Based Assessment of ADHD in Youth Using a Receiver Operating Characteristic Approach. *Journal of Clinical Child & Adolescent Psychology, 47*(5), 808-820. doi:10.1080/15374416.2016.1225502

Jiménez-Figueroa, G., Ardila-Duarte, C., Pineda, D. A., Acosta-López, J. E., Cervantes-Henríquez, M. L., Pineda-Alhucema, W., . . . Vélez, J. I. (2017). Prepotent response inhibition and reaction times in children with attention deficit/hyperactivity disorder from a Caribbean community. *ADHD Attention deficit and hyperactivity disorders, 9*(4), 199-211.

Johnson, K. A., Kelly, S. P., Robertson, I. H., Barry, E., Mulligan, A., Daly, M., . . . Hawi, Z. (2008). Absence of the 7-repeat variant of the DRD4 VNTR is associated with drifting sustained attention in children with ADHD but not in controls. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 147*(6), 927-937.

Jonas, K. G., & Markon, K. E. (2014). A Meta-Analytic Evaluation of the Endophenotype Hypothesis: Effects of Measurement Paradigm in the Psychiatric Genetics of Impulsivity. *J Abnorm Psychol, 123*(3), 660-675. doi:10.1037/a0037094

Kail, R. (1992). Processing speed, speech rate, and memory. *Dev Psychol, 28*(5), 904.

Kail, R., & Park, Y. (1994). Processing time, articulation time, and memory span. *J Exp Child Psychol, 57*, 291.

Kail, R., & Salthouse, T. A. (1994). Processing speed as a mental capacity. *Acta Psychol (Amst), 86*, 225.

Kalff, A. C., De Sonneville, L. M., Hurks, P. P. M., Hendriksen, J. G. M., Kroes, M., Feron, F. J. M., . . . Jolles, J. (2005). Speed, speed variability, and accuracy of information processing in 5 to 6-year-old children at risk of ADHD. *Journal of the International Neuropsychological Society, 11*, 183.

Karalunas, S. L., Bierman, K. L., & Huang-Pollock, C. L. (2016). Test–retest reliability and measurement invariance of executive function tasks in young children with and without ADHD. *Journal of Attention Disorders*, 1087054715627488.

Karalunas, S. L., Gustafsson, H. C., Dieckmann, N. F., Tipsord, J., Mitchell, S. H., & Nigg, J. T. (2017). Heterogeneity in Development of Aspects of Working Memory Predicts Longitudinal Attention Deficit Hyperactivity Disorder Symptom Change. *J Abnorm Psychol, 126*(6), 774-792. doi:10.1037/abn0000292

Karalunas, S. L., & Huang-Pollock, C. L. (2013). Integrating impairments in reaction time and executive function using a diffusion model framework. *J Abnorm Child Psychol, 41*(5), 837-850. doi:10.1007/s10802-013-9715-2

Karalunas, S. L., Huang-Pollock, C. L., & Nigg, J. T. (2012). Decomposing Attention-Deficit/Hyperactivity Disorder (ADHD)-related effects in response speed and variability. *Neuropsychology, 26*(6), 684-694. doi:10.1037/a0029936

Kebir, O., & Joober, R. (2011). Neuropsychological endophenotypes in attention-deficit/hyperactivity disorder: a review of genetic association studies. *European Archives of Psychiatry and Clinical Neuroscience, 261*(8), 583-594. doi:10.1007/s00406-011-0207-5

Kebir, O., Tabbane, K., Sengupta, S., & Joober, R. (2009). Candidate genes and neuropsychological phenotypes in children with ADHD: review of association studies. *Journal of Psychiatry & Neuroscience, 34*(2), 88-101.

Kessler, R. C., Adler, L. A., Gruber, M. J., Sarawate, C. A., Spencer, T., & Van Brunt, D. L. (2007). Validity of the World Health Organization Adult ADHD Self-Report Scale (ASRS) Screener in a representative sample of health plan members. *International Journal of Methods in Psychiatric Research, 16*(2), 52-65.

Kieling, C., Roman, T., Doyle, A. E., Hutz, M. H., & Rohde, L. A. (2006). Association between DRD4 gene and performance of children with ADHD in a test of sustained attention. *Biological Psychiatry, 60*(10), 1163-1165.

Kofler, M. J., Rapport, M. D., Sarver, D. E., Raiker, J. S., Orban, S. A., Friedman, L. M., & Kolomeyer, E. G. (2013). Reaction time variability in ADHD: A meta-analytic review of 319 studies. *Clinical Psychology Review, 33*(6), 795-811. doi:10.1016/j.cpr.2013.06.001

Kraemer, H. C. (1992). *Evaluating medical tests: objective and quantitiative guidelines* (Vol. 26). Newbury Park, CA: Sage Publications.

Kuntsi, J., Neale, B. M., Chen, W., Faraone, S. V., & Asherson, P. (2006). The IMAGE project: methodological issues for the molecular genetic analysis of ADHD. *Behavioral and Brain Functions, 2*(1), 27.

Kuntsi, J., Oosterlaan, J., & Stevenson, J. (2001). Psychological mechanisms in hyperactivity: I response inhibition deficit, working memory impairment, delay aversion, or something else? *Journal of child Psychology and Psychiatry, 42*(2), 199-210. doi:10.1111/1469-7610.00711

Kuntsi, J., Rogers, H., Swinard, G., Borger, N., van der Meere, J., Rijsdijk, F., & Asherson, P. (2006). Reaction time, inhibition, working memory and 'delay aversion' performance: genetic influences and their interpretation. *Psychol Med, 36*(11), 1613-1624. doi:10.1017/S0033291706008580

Kuntsi, J., & Stevenson, J. (2001). Psychological mechanisms in hyperactivity: II The role of genetic factors. *Journal of child Psychology and Psychiatry, 42*(2), 211-219. doi:10.1111/1469-7610.00712

Lacouture, Y., & Cousineau, D. (2008). How to use MATLAB to fit the ex-Gaussian and other probability functions to a distribution of response times. *Tutorials in Quantitative Methods for Psychology, 4*(1), 35-45.

Lahey, B. B., Applegate, B., McBurnett, K., Biederman, J., Greenhill, L., Hynd, G. W., . . . Shaffer, D. (1994). DSM-IV field trials for Attention-Deficit Hyperactivity Disorder in Children and Adolescents. *American Journal of Psychiatry, 151*(11), 1673-1685.

Lampert, T. L., Polanczyk, G., Tramontina, S., Mardini, V., & Rohde, L. A. (2004). Diagnostic performance of the CBCL-Attention Problem Scale as a screening measure in a sample of Brazilian children with ADHD. *Journal of Attention Disorders, 8*(2), 63-71. doi:10.1177/108705470400800204

Larson, G. E., & Alderton, D. L. (1990). Reaction time variability and intelligence: A "worst performance" analysis of individual differences. *Intelligence, 14*(3), 309-325. doi:http://dx.doi.org/10.1016/0160-2896(90)90021-K

Lee, S. S., & Hinshaw, S. P. (2006). Predictors of adolescent functioning in girls with attention deficit hyperactivity disorder (ADHD): the role of childhood ADHD, conduct problems, and peer status. *Journal of Clinical Child and Adolescent Psychology, 35*(3), 356-368.

Lenzenweger, M. F. (2013). ENDOPHENOTYPE, INTERMEDIATE PHENOTYPE, BIOMARKER: DEFINITIONS, CONCEPT COMPARISONS, CLARIFICATIONS. *Depression and anxiety, 30*(3), 185-189. doi:10.1002/da.22042

Leth-Steensen, C., Elbaz, Z. K., & Douglas, V. I. (2000). Mean response times, variability, and skew in the responding of ADHD children: a response time distributional approach. *Acta Psychol (Amst), 104*, 190.

Lijffijt, M., Kenemans, J. L., Verbaten, M. N., & van Engeland, H. (2005). A meta-analytic review of stopping performance in attention-deficit/hyperactivity disorder: deficient inhibitory motor control? *J Abnorm Psychol, 114*(2), 216-222. doi:10.1037/0021-843X.114.2.216

Logan, G. D. (1985). On the ability to inhibit simple thoughts and actions: II. Stop-signal studies of repetition priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*(4), 691.

Luce, R. D. (1986). *Response Times*. New York: Oxford University Press.

Madden, D. J., Gottlob, L. R., Denny, L. L., Turkington, T. G., Provenzale, J. M., Hawk, T. C., & Coleman, R. E. (1999). Aging and recognition memory: Changes in regional cerebral blood flow associated with components of reaction time distributions. *Journal of Cognitive Neuroscience, 11*(5), 511-520. doi:10.1162/089892999563571

Magimairaj, B. M., & Montgomery, J. W. (2012). Children's verbal working memory: relative importance of storage, general processing speed, and domain-general controlled attention. *Acta Psychol (Amst), 140*(3), 196-207. doi:10.1016/j.actpsy.2012.05.004

Martinussen, R., Hayden, J., Hogg-Johnson, S., & and Tannock, R. (2005). A meta-analysis of working memory impairments in children with attention-deficit/hyperactivity disorder. *Journal of the American Academy of Child and Adolescent Psychiatry, 44*(4), 384.

Matzke, D., & Wagenmakers, E. J. (2009). Psychological interpretation of the ex-Gaussian and shifted Wald parameters: a diffusion model analysis. *Psychon Bull Rev, 16*(5), 798-817. doi:10.3758/PBR.16.5.798

Mealer, C., Morgan, S., & Luscomb, R. (1996). Cognitive functioning of ADHD and non-ADHD boys on the WISC-III and WRAML: An analysis within a memory model. *Journal of Attention Disorders, 1*(3), 133-145.

Mella, N., Fagot, D., Lecerf, T., & de Ribaupierre, A. (2015). Working memory and intraindividual variability in processing speed: A lifespan developmental and individual-differences study. *Memory & Cognition, 43*(3), 340-356. doi:10.3758/s13421-014-0491-1

Minde, K., Eakin, L., Hechtman, L., Ochs, E., Bouffard, R., Greenfield, B., & Looper, K. (2003). The psychosocial functioning of children and spouses of adults with ADHD. *Journal of*

Child Psychology and Psychiatry and Allied Disciplines, 44(4), 637-646. doi:10.1111/1469-7610.00150

Moret-Tatay, C., Leth-Steensen, C., Irigaray, T. Q., Argimon, I. I. L., Gamermann, D., Abad-Tortosa, D., . . . Castella, P. (2016). The Effect of Corrective Feedback on Performance in Basic Cognitive Tasks: An Analysis of RT Components. *Psychologica Belgica, 56*(4), 370-381. doi:10.5334/pb.240

Myerson, J., Hale, S., Zheng, Y., Jenkins, L., & Widaman, K. F. (2003). The difference engine: A model of diversity in speeded cognition. *Psychonomic Bulletin & Review, 10*(2), 262-288.

Nigg, J. T. (1999). The ADHD response-inhibition deficit as measured by the stop task: Replication with DSM-IV combined type, extension, and qualification. *J Abnorm Child Psychol, 27*(5), 393-402. doi:10.1023/a:1021980002473

Nigg, J. T., & Casey, B. J. (2005). An integrative theory of attention-deficit/hyperactivity disorder based on the cognitive and affective neurosciences. *Development and Psychopathology, 17*(3), 785-806. doi:10.1017/s0954579405050376

Nigg, J. T., Gustafsson, H. C., Karalunas, S. L., Ryabinin, P., McWeeney, S. K., Faraone, S. V., . . . Wilmot, B. (2018). Working Memory and Vigilance as Multivariate Endophenotypes Related to Common Genetic Risk for Attention-Deficit/Hyperactivity Disorder. *Journal of the American Academy of Child and Adolescent Psychiatry, 57*(3), 175-182. doi:10.1016/j.jaac.2017.12.013

Nigg, J. T., Nikolas, M., & Burt, S. A. (2010). Measured Gene-by-Environment Interaction in Relation to Attention-Deficit/Hyperactivity Disorder. *Journal of the American Academy of Child and Adolescent Psychiatry, 49*(9), 863-873. doi:10.1016/j.jaac.2010.01.025

Nigg, J. T., Nikolas, M., Knottnerus, G. M., Cavanagh, K., & Friderici, K. (2010). Confirmation and extension of association of blood lead with attention-deficit/hyperactivity disorder (ADHD) and ADHD symptom domains at population-typical exposure levels. *Journal of child Psychology and Psychiatry, 51*(1), 58-65. doi:10.1111/j.1469-7610.2009.02135.x

Nigg, J. T., Willcutt, E. G., Doyle, A. E., & Sonuga-Barke, E. J. (2005). Causal heterogeneity in attention-deficit/hyperactivity disorder: do we need neuropsychologically impaired subtypes? *Biol Psychiatry, 57*(11), 1224-1230. doi:10.1016/j.biopsych.2004.08.025

Oosterlaan, J., Logan, G. D., & Sergeant, J. A. (1998). Response Inhibition in AD/HD, CD, Comorbid AD/HD + CD, Anxious, and Control Children: A Meta-analysis of Studies with the Stop Task. *J. Child Psychait., 29*(3), 425.

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. doi:10.1016/j.cpr.2011.02.005

Pauli-Pott, U., Dalir, S., Mingebach, T., Roller, A., & Becker, K. (2013). Do different ADHD-related etiological risks involve specific neuropsychological pathways? An analysis of mediation processes by inhibitory control and delay aversion. *Journal of child Psychology and Psychiatry, 54*(7), 800-809. doi:10.1111/jcpp.12059

Pelham, W. E., Foster, E. M., & Robb, J. A. (2007). The economic impact of Attention-Deficit/Hyperactivity disorder in children and adolescents. *Journal of Pediatric Psychology, 32*(6), 711-727. doi:10.1093/jpepsy/jsm022

Pineda, D. A., Puerta, I. C., Aguirre, D. C., García-Barrera, M. A., & Kamphaus, R. W. (2007). The role of neuropsychologic tests in the diagnosis of attention deficit hyperactivity disorder. *Pediatric neurology, 36*(6), 373-381.

Pinto, R., Asherson, P., Ilott, N., Cheung, C. H. M., & Kuntsi, J. (2016). Testing for the Mediating Role of Endophenotypes Using Molecular Genetic Data in a Twin Study of ADHD Traits. *American Journal of Medical Genetics Part B-Neuropsychiatric Genetics, 171*(7), 982-992. doi:10.1002/ajmg.b.32463

Raiker, J. S., Freeman, A. J., Perez-Algorta, G., Frazier, T. W., Findling, R. L., & Youngstrom, E. A. (2017). Accuracy of Achenbach Scales in the Screening of Attention-Deficit/Hyperactivity Disorder in a Community Mental Health Clinic. *Journal of the American Academy of Child and Adolescent Psychiatry, 56*(5), 401-409. doi:10.1016/j.jaac.2017.02.007

Rasmussen, K., & Levander, S. (2009). Untreated ADHD in adults: are there sex differences in symptoms, comorbidity, and impairment? *Journal of Attention Disorders, 12*(4), 353-360.

Ratcliff, R. (2002). A diffusion model account of response time and accuracy in a brightness discrimination task: Fitting real data and failing to fit fake but plausible data. *Psychonomic Bulletin & Review, 9*(2), 278-291. doi:10.3758/bf03196283

Ratcliff, R. (2014). Measuring psychometric functions with the diffusion model. *Journal of Experimental Psychology: Human Perception and Performance, 40*(2), 870.

Ratcliff, R., Love, J., Thompson, C. A., & Opfer, J. E. (2012). Children are not like older adults: A diffusion model analysis of developmental changes in speeded responses. *Child Development, 83*(1), 367-381.

Ratcliff, R., & Murdock, B. B. (1976). Retrieval processes in recognition memory. *Psychological Review, 83*(3), 190.

Ratcliff, R., & Smith, P. L. (2010). Perceptual discrimination in static and dynamic noise: the temporal relation between perceptual encoding and decision making. *Journal of Experimental Psychology: General, 139*(1), 70.

Reimers, S., & Maylor, E. A. (2006). Gender effects on reaction time variability and trial-to-trial performance: reply to Deary and Der (2005). *Aging, Neuropsychology, and Cognition, 13*(3-4), 479-489.

Rey, J., Morris-Yates, A., & Stanislaw, H. (1992). Measuring the accuracy of diagnostic tests using receiver operating characteristics (ROC) analysis. *International Journal of Methods in Psychiatric Research, 2*, 39-50.

Reynolds, C. R., & Horton, A. M. (2014). The Neuropsychology of Executive Functioning and the DSM-5. In S. Goldstein & J. A. Naglieri (Eds.), *Handbook of Executive Functioning* (pp. 89-105). New York, NY: Springer New York.

Reynolds, C. R., & Kamphaus, R. W. (2004). *Behavior Assessment for Children, (BASC-2).* Circle Pines, MN: American Guidance Service.

Roid, G. H. (2003). *Stanford Binet intelligence scales.* Itasca, IL: Riverside Publications.

Rotello, C. M., & Zeng, M. (2008). Analysis of RT distributions in the remember-know paradigm. *Psychonomic Bulletin & Review, 15*(4), 825-832. doi:10.3758/pbr.15.4.825

Sadeghi-Bazargani, H., Amiri, S., Hamraz, S., Malek, A., Abdi, S., & Shahrokhi, H. (2014). Validity and reliability of the Persian version of Conner's adult ADHD rating scales: observer and self-report screening versions. *Journal of Clinical Research & Governance, 3*(1), 42-47.

Salthouse, T. A. (1991). Mediation of Adult Age Differences in Cognition by Reductions in Working Memory and Speed Processing. *Psychological Science, 2*(3), 179-189.

Salthouse, T. A. (1996). The Processing-Speed Theory of Adult Age Difference in Cognition. *Psychological Review, 103*(3), 403-428.

Sattler, J. (2008). *Resource Guide to Accompany Assessment of Children: Cognitive Foundations, 5th Edition*. San Diego: Jerome Sattler Publisher, Inc.

Schachar, R. J., Chen, S., Logan, G. D., Ornstein, T. J., Crosbie, J., Ickowicz, A., & Pakulak, A. (2004). Evidence for an error monitoring deficit in attention deficit hyperactivity disorder. *J Abnorm Child Psychol, 32*(3), 285-293. doi:10.1023/B:JACP.0000026142.11217.f2

Scheres, A., Tontsch, C., Thoeny, A. L., & Kaczkurkin, A. (2010). Temporal Reward Discounting in Attention-Deficit/Hyperactivity Disorder: The Contribution of Symptom Domains, Reward Magnitude, and Session Length. *Biological Psychiatry, 67*(7), 641-648. doi:10.1016/j.biopsych.2009.10.033

Schmiedek, F., Oberauer, K., Wilhelm, O., Suss, H. M., & Wittmann, W. W. (2007). Individual differences in components of reaction time distributions and their relations to working memory and intelligence. *J Exp Psychol Gen, 136*(3), 414-429. doi:10.1037/0096-3445.136.3.414

Schwartz, K., & Verhaeghen, P. (2008). ADHD and Stroop interference from age 9 to age 41 years: a meta-analysis of developmental effects. *Psychological Medicine, 38*(11), 1607-1616.

Shaffer, D., Fisher, P., Lucas, C. P., Dulcan, M. K., & Schwab-Stone, M. E. (2000). NIMH Diagnostic Interview Schedule for Children Version IV (NIMH DISC-IV). *Journal of the American Academy of Child & Adolescent Psychiatry, 39*(1), 28-38. doi:10.1097/00004583-200001000-00014

Shahar, N., Teodorescu, A. R., Usher, M., Pereg, M., & Meiran, N. (2014). Selective Influence of Working Memory Load on Exceptionally Slow Reaction Times. *Journal of Experimental Psychology-General, 143*(5), 1837-1860. doi:10.1037/a0037190

Sheskin, D. J. (2003). *Handbook of parametric and nonparametric statistical procedures*: crc Press.

Smith, & Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. *Trends Neurosci, 27*(3), 161-168. doi:10.1016/j.tins.2004.01.006

Smith, C. J. (2012). Diagnostic tests (1) - sensitivity and specificity. *Phlebology, 27*(5), 250-251. doi:10.1258/phleb.2012.012J05

Snow, J., & Hynd, G. W. (1984). Determining neuropsychological" strengths" and" weaknesses" on the Luria-Nebraska: Good practice or wishful thinking?

Solanto, M. V., Wender, E. H., & Bartell, S. S. (1997). Effects of methylphenidate and behavioral contingencies on sustained attention in attention-deficit hyperactivity disorder: a test of the reward dysfunction hypothesis. *Journal of Child and Adolescent Psychopharmacology, 7*(2), 123-136.

Swets, J. A. (1982). SENSITIVITIES AND SPECIFICITIES OF DIAGNOSTIC-TESTS. *Jama-Journal of the American Medical Association, 248*(5), 548-549. doi:10.1001/jama.248.5.548

Szomlaiski, N., Dyrborg, J., Rasmussen, H., Schumann, T., Koch, S., & Bilenberg, N. (2009). Validity and clinical feasibility of the ADHD rating scale (ADHD-RS) A Danish Nationwide Multicenter Study. *Acta Paediatrica, 98*(2), 397-402.

Teicher, M. H., Polcari, A., Fourligas, N., Vitaliano, G., & Navalta, C. P. (2012). Hyperactivity persists in male and female adults with ADHD and remains a highly discriminative

feature of the disorder: a case-control study. *Bmc Psychiatry, 12*, 14. doi:10.1186/1471-244x-12-190

Thissen, A., Rommelse, N. N. J., Hoekstra, P. J., Hartman, C., Heslenfeld, D., Luman, M., . . . Buitelaar, J. K. (2014). Attention deficit hyperactivity disorder (ADHD) and executive functioning in affected and unaffected adolescents and their parents: challenging the endophenotype construct. *Psychological Medicine, 44*(4), 881-892. doi:10.1017/s0033291713001153

Tye, C., Johnson, K. A., Kelly, S. P., Asherson, P., Kuntsi, J., Ashwood, K. L., . . . McLoughlin, G. (2016). Response time variability under slow and fast-incentive conditions in children with ASD, ADHD and ASD plus ADHD. *Journal of child Psychology and Psychiatry, 57*(12), 1414-1423. doi:10.1111/jcpp.12608

Unsworth, N., Redick, T. S., Lakey, C. E., & Young, D. L. (2010). Lapses in sustained attention and their relation to executive control and fluid abilities: An individual differences investigation. *Intelligence, 38*(1), 111-122. doi:10.1016/j.intell.2009.08.002

van Belle, J., van Hulst, B. M., & Durston, S. (2015). Developmental differences in intra-individual variability in children with ADHD and ASD. *Journal of child Psychology and Psychiatry, 56*(12), 1316-1326. doi:10.1111/jcpp.12417

Van Zandt, T., & Townsend, J. T. (2014). Designs for and analyses of response time experiments. *The Oxford Handbook of Quantitative Methods: Foundations, 1*, 260.

Vicari, S., Caravale, B., Carlesimo, G. A., Casadei, A. M., & Allemand, F. (2004). Spatial working memory deficits in children at ages 3-4 who were low birth weight, preterm infants. *Neuropsychology, 18*(4), 673-678. doi:10.1037/0894-4105.18.4.673

Volkow, N. D., Wang, G.-J., Newcorn, J. H., Kollins, S. H., Wigal, T. L., Telang, F., . . . Logan, J. (2011). Motivation deficit in ADHD is associated with dysfunction of the dopamine reward pathway. *Molecular Psychiatry, 16*(11), 1147.

Voss, A., Nagler, M., & Lerche, V. (2013). Diffusion Models in Experimental Psychology A Practical Introduction. *Experimental Psychology, 60*(6), 385-402. doi:10.1027/1618-3169/a000218

Voss, A., Rothermund, K., & Voss, J. (2004). Interpreting the parameters of the diffusion model: An empirical validation. *Memory & Cognition, 32*(7), 1206-1220.

Wahlstedt, C., Thorell, L. B., & Bohlin, G. (2009). Heterogeneity in ADHD: Neuropsychological Pathways, Comorbidity and Symptom Domains. *J Abnorm Child Psychol, 37*(4), 551-564. doi:10.1007/s10802-008-9286-9

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children, 4th Ed (WISC-IV) Technical and Interpretive Manual*. San Antonio: Harcourt Brace.

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale- Fourth Edition (WAIS-IV)*. San Antonio, TX: NCS Pearson.

Weigard, A., & Huang-Pollock, C. L. (2016). The role of speed in ADHD-related working memory deficits: A time-based resource-sharing and diffusion model account. *Clinical Psychological Science, 5*(2), 195-211.

White, C. N., Ratcliff, R., & Starns, J. J. (2011). Diffusion models of the flanker task: Discrete versus gradual attentional selection. *Cognitive Psychology, 63*(4), 210-238. doi:10.1016/j.cogpsych.2011.08.001

Wiener, J., Malone, M., Varma, A., Markel, C., Biondic, D., Tannock, R., & Humphries, T. (2012). Children's perceptions of their ADHD symptoms: Positive illusions, attributions, and stigma. *Canadian Journal of School Psychology, 27*(3), 217-242.

Wilens, T. E., Biederman, J., Faraone, S. V., Martelon, M., Westerberg, D., & Spencer, T. J. (2009). Presenting ADHD symptoms, subtypes, and comorbid disorders in clinically referred adults with ADHD. *The Journal of clinical psychiatry, 70*(11), 1557.

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. *Biol Psychiatry, 57*(11), 1336-1346. doi:10.1016/j.biopsych.2005.02.006

Willcutt, E. G., Pennington, B. F., Olson, R. K., Chhabildas, N., & Hulslander, J. (2005). Neuropsychological analyses of comorbidity between reading disability and attention deficit hyperactivity disorder: in search of the common deficit. *Dev Neuropsychol, 27*(1), 35-78. doi:10.1207/s15326942dn2701_3

Wilson, V. B., Mitchell, S. H., Musser, E. D., Schmitt, C. F., & Nigg, J. T. (2011). Delay discounting of reward in ADHD: application in young children. *Journal of child Psychology and Psychiatry, 52*(3), 256-264.

Yin, J. J., & Tian, L. L. (2014). Joint confidence region estimation for area under ROC curve and Youden index. *Statistics in Medicine, 33*(6), 985-1000. doi:10.1002/sim.5992

Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer, 3*(1), 32-35.

Youngstrom, E. A. (2014). A Primer on Receiver Operating Characteristic Analysis and Diagnostic Efficiency Statistics for Pediatric Psychology: We Are Ready to ROC. *Journal of Pediatric Psychology, 39*(2), 204-221. doi:10.1093/jpepsy/jst06

Hilary Galloway-Long, M.S.
Vita

Education

2020 **PhD: Clinical Psychology**
The Pennsylvania State University: University Park, PA

2016 **MS: Clinical Psychology**
The Pennsylvania State University: University Park, PA

2010 **BA: Psychology**
Lewis and Clark College: Portland, OR

Publications

Galloway-Long, H., Huang-Pollock, C., Bierman, K., Neely, K. (*In preparation*). Ahead of the (ROC) curve: A Statistical approach to utilizing ex-Gaussian parameters of reaction time in diagnosing ADHD across three developmental periods.

Weigard, A., Wilson, S., Shapiro, Z., **Galloway-Long, H.,** Huang-Pollock, C. (*Under Review*). Neural correlates of working memory load's suppression of aversive olfactory distraction effects. *Brain Imaging and Behavior.*

Huang-Pollock, C., Shapiro, Z., **Galloway-Long, H.,** & Feldman, J. (In Press). A practical guide for designing and conducting experimental studies in child psychopathology. In Aidan G. Wright & Michael N. Hallquist (Eds.), *Handbook of Research Methods in Clinical Psychology.* Cambridge University Press.

Galloway-Long, H., & Huang-Pollock, C. (2018). Using inspection time and ex-Gaussian parameters of reaction time to predict executive functions in children with ADHD. *Intelligence, 69, 186-194.*

Huang-Pollock, C., Shapiro, Z., **Galloway-Long, H., &** Weigard, A., (2017). Is poor working memory a transdiagnostic risk factor for psychopathology? *Journal of Abnormal Child Psychology, 45, 1477-1490.*

Galloway-Long, H., Shapiro, Z., & Huang-Pollock, C. (2016). Diffusion modeling in ADHD: A brief introduction and application for clinical practice. *National Association for Neuropsychology Bulletin, 30, 19-21.*

Huang-Pollock, C., Ratcliff, R., McKoon, G., Shapiro, Z., Weigard, A., & **Galloway-Long, H.** (2016). Using the diffusion decision model to explain cognitive deficits in Attention-Deficit Hyperactivity Disorder. *Journal of Abnormal Child Psychology, 45, 57-68.*

Nigg, J. T., Johnstone, J. M., Musser, E. D., **Galloway-Long, H.,** Willoughby, M. T., & Shannon, J. (2016). Attention-deficit/hyperactivity disorder (ADHD) and being overweight/obesity: New data and meta-analysis. *Clinical Psychology Review, 43, 67-79.*

Musser, E.D., **Galloway-Long, H.S.**, Frick, P.J., Nigg, J.T. (2013). Autonomic subtypes of attention-deficit/hyperactivity disorder in children, *Journal of the American Academy of Child and Adolescent Psychiatry 52, 163-171.*