Sex Differences in the Executive Function Profile of Attention-Deficit/Hyperactivity Disorder: A Meta-Analytic Review

A Thesis Presented for the
Master of Arts
Degree
The University of Tennessee, Knoxville

Jenna Elizabeth Gilmore
August 2018

Acknowledgments

I am immensely grateful to my research mentor and thesis chair, Dr. Jennifer Bolden, for her encouragement and direction. Her passion for “egghead” research questions and commitment to working with children and families with ADHD are inspiring. I would also like to thank my thesis committee members, Drs. Aaron Buss and Todd Moore, whose expertise in the development of executive functions and meta-analyses, respectively, strengthened my work. Additionally, I would like to thank the many undergraduate research students who lent me their time and attention to detail, as well as my graduate research colleagues, J. Parks Fillauer and Emily Mariotti. Thank you to my parents, Dr. Susan Benson and Mark Gilmore, for emboldening me to think critically and to be curious. Finally, I am indebted to my partner, Ronnie Matharu, for his patience, support, and practicality.

Abstract

Attention-Deficit/Hyperactivity Disorder (ADHD) is defined as a neurodevelopmental disorder, and literature examining executive function (EF) impairments within ADHD samples continues to grow. Moreover, much work has been done to promote investigation of sex differences in ADHD given that the ratio of boys to girls diagnosed with ADHD. It is thus surprising relatively little is known about the specific executive function profiles of girls and boys with ADHD. The current study provides meta-analytic and qualitative summaries of 22 studies of ADHD EF profiles for girls and boys with ADHD. Analyses were separated according to five proposed domains of EF. Results yielded no significant differences between sexes for Attentional Control (d = -0.071, p = 0.417), Inhibition (d = -0.102, p = 0.319), Set-Shifting (d = -0.19, p = .16), Planning/Organization (d = 0.009, p = 0.944), and Phonological Working Memory (d = 0.146, p = 0.210). Collectively, results suggest similar EF profiles for girls and boys diagnosed with ADHD. However, given that only 22 studies met inclusion criteria for quantitative synthesis, more work is needed. Recommendations for future research are provided.

Table of Contents

Chapter 1 Introduction 1
Sex Differences in the Prevalence of ADHD 1
Sex-Specific Syomptomatology of ADHD 2
The Known Profile of Girls with ADHD 3
Brain Development and ADHD 4
Models of ADHD and Executive Function 6
Existing Literature on ADHD and Executive Function 7
Funding Concerns for Current ADHD Research 9
Study Aims 10
Chapter 2 Methods 11
Literature Search and Inclusion Criteria 11
Identification of Studies 12
Dependent Variables 13
Included Studies 16
Chapter 3 Analyses 18
Calculating Effect Sizes 18
Assumptions of Meta-Analyses 18
Meta-Regression 19
Chapter 4 Results 21
Analysis of Effect Size Distribution for Attentional Control 21
Analysis of Effect Size Distribution for Inhibition 21
Analysis of Effect Size Distribution for Set-Shifting 21
Analysis of Effect Size Distribution for Planning/Organization 22
Analysis of Effect Size Distribution for Phonological Working Memory 22
Chapter 5 Discussion 24
Summary 24
Conclusions 27
Limitations 28
Future Directions 30
References 33
Appendices 42
Appendix A Tables 43
Appendix B Figures 52
Appendix C Codebook Manual 55
Vita 58

List of Tables

Table 1: *Effect Sizes and Moderators for Attentional Control* 44
Table 2: *Effect Sizes and Moderators for Inhibition*.....46
Table 3. *Effect Sizes and Moderators for Planning/Organization*.....48
Table 4. *Effect Sizes and Moderators for Set-Shifting*.....50
Table 5. *Effect Sizes and Moderators for Phonological Working Memory*.....51

List of Figures

Figure 1: *Flow Chart for Included Studies* 53

Chapter 1

Introduction

Sex Differences in the Prevalence of ADHD

Attention-deficit/hyperactivity disorder (ADHD) is a neurodevelopmental disorder marked by symptoms of inattention, hyperactivity, and impulsivity, and impairments are pervasive across environments. The prevalence of pediatric ADHD in the U.S. ranges between 3-7% (Polanczyk, de Lima, Horta, Biederman, & Rohde, 2007). The *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.; *DSM-5*; American Psychiatric Association, 2013) reports the ratio of males to females with ADHD as 2:1 for children in the community and 6:1 in clinical samples. While the male to female ratios for pediatric ADHD are unequal in both clinical and community settings, the gap is much narrower in a community sample. It is important to further clarify the role of setting in explicating discrepancies between clinic and community ratios, as well as to understand the convergence and divergence of ADHD manifestations in girls and boys with ADHD. Children presenting in clinical settings (e.g., outpatient psychology clinic) have either already been diagnosed with ADHD or have been identified as having attention, hyperactivity, and/or impulsivity concerns and are presenting to the clinic for intervention and/or diagnostic clarity. Community samples, however, do not require such identifications prior to participating in an ADHD study. Regardless of the sample (i.e., community or clinic), a number of questions arise regarding the male to female ratio. For example, are girls with ADHD typically less impaired than boys with ADHD and, thus, need less intensive or different interventions? Alternatively, is the typical manifestation of ADHD in a female population characteristically different and, consequently, difficult to identify by traditional means?

Sex-Specific Symptomatology of ADHD

One potential explanation for the discrepancy between community and clinical samples is the role teachers play in identifying ADHD symptomatology and referring children to clinical settings for further evaluation. Research, including that from meta-analytic reviews, finds teachers are more likely to endorse ADHD symptomatology in boys than in girls (Gaub & Carlson, 1997; Gershon, 2002; Sciutto, Nolfi, & Bluhm, 2004). Sciutto et al. (2004) provided teachers with a fictional profile of a child's academic record and were asked whether they would refer the fictional child for a clinical evaluation. Profiles varied by gender and by symptom presentation (inattention, hyperactivity, or hyperactivity with aggression). Teachers were more likely to refer male profiles than female profiles for evaluation, regardless of symptom presentation; however, the hyperactive-only symptom profile evidenced the largest gender bias. Abikoff (1991) compared the observed gender-specific classroom behavior norms of children with ADHD and found girls with ADHD engaged in more verbal aggression than same-sex peers, while boys with ADHD engaged in more rule-breaking and externalizing behaviors. Taken together, these studies indicate teacher endorsement of ADHD may be influenced by teachers' expectations for gender differences in ADHD symptoms. However, it also appears girls and boys present different behavioral profiles of ADHD symptomatology, about which teachers may be uninformed. This is consistent with research suggesting the use of sex-specific or gender-specific norms for ADHD diagnosis (Mahone & Wodka, 2008; Waschbusch & King, 2006). For example, Waschbusch and King (2006) identified a subset of girls with higher levels of ADHD and ODD symptoms compared to their same-sex peers. Although the girls did not meet criteria for either disorder, they were nearly as impaired on either of two measures, the Assessment of Disruptive Symptoms – DSM-IV Version (ADS-IV) or the Children's Impairment Rating Scale

(CIRS), as were girls meeting *DSM-IV* criteria for ADHD. Therefore, more work is needed in clarifying male and female profiles of ADHD and in disseminating these profiles to teachers and other potential referral sources.

The Known Profile of Girls With ADHD

To date, two meta-analytic reviews examine social, academic, and intellectual functioning of girls and boys with ADHD (Gaub & Carlson, 1997; Gershon, 2002). Neither review documented gender differences in academic performance of children with ADHD. Gaub and Carlson (1997) and Gershon (2002) both included mathematics (d = -08 to .03), reading (d = .10 to .11), spelling (d =.14 to .22) as academic variables. Gaub and Carlson (1997) also provided a comparison of academic language performance (d = -0.10). There were no significant gender differences for the following social variables: peer liking [d = 0.16; (Gaub & Carlson, 1997)], peer disliking [d = .10; (Gaub & Carlson, 1997)], peer popularity [d = .03; (Gershon, 2002)], social skills [d = -.23; (Gershon, 2002)], or social problems [d = .08; (Gershon, 2002)]. Gaub and Carlson (1997) found that boys with ADHD exhibit more peer aggression than do girls with ADHD (d = 0.35) . Comorbidity ratings across meta-analyses emerged as one way the profiles of girls and boys with ADHD might diverge. Gershon (2002) found girls with ADHD were more likely to be rated by parents and teachers as exhibiting comorbid internalizing conditions (d = -.12). Both studies found boys were more likely to be rated as exhibiting comorbid externalizing conditions. Gaub and Carlson (1997) divided externalizing conditions by Conduct Disorder (d = 0.14) and Other Externalizing (d =.17), while Gershon (2002) provided a single, general measure of externalizing comorbidity (Externalizing; d =.21). Notably, Gershon posited his own meta-analyses might be comparing apples to oranges, such that ADHD girls with comorbid anxiety and/or mood problems were compared to ADHD boys with comorbid conduct

and oppositional defiant disorders. Divergent comorbidity profiles between girls and boys with ADHD present significant challenges to gender analyses in any study, as well as to analyses combining genders. However, although Gershon (2002) found no differences between parent and teacher ratings of externalizing behaviors, Waschbusch and King (2006) shed important light on the vagueness of assessment directions. Specifically, while one rater may correctly compare a girl to other girls only, another rater may compare a girl to both girls *and boys* (Waschbusch & King, 2006). As this can be a potential pitfall for both parents and teachers, meta-analysis of differences between boys and girls with ADHD on objective assessments (e.g., neuropsychological functioning) may be more useful.

Brain Development and ADHD

Extant research on normal brain development highlights a sexual dimorphism, whereby girls mature anywhere from one to three years earlier than do boys (Mahone & Wodka, 2008) (Asato, Terwilliger, Woo, & Luna, 2010; Lenroot et al., 2007; Razanhan et al., 2011; Ruigrok et al., 2014). It is important for future research to distinguish between structural and functional sex differences. While investigations of the brain structure of typically developing brains routinely find sexual dimorphism across development, results from functional investigations are less clear. Overall, there is some support for null effects of sex by executive function domain within the developmental literature. Set-shifting, in particular, does not appear to differ across sexes (Gathercole, Pickering, Ambridge, and Wearing, 2004; Huizinga & van der Molen, 2007). Huizinga and van Der Molen (2007) studied performance of a cross-section of typically developing children on the Wisconsin Card Sorting Test (WCST). Comparison of age groups suggested set-shifting and inhibition mature around 11 years of age but working memory performance continued to develop into adolescence. No sex differences were observed. Multi-

method assessment of working memory in children and adolescents from 4-15 years of age yielded no broad sex differences; however adolescent boys outperformed adolescent girls on the Visual Patterns Test and block recall (Gathercole, Pickering, Ambridge, and Wearing, 2004).

Within the ADHD literature, meta-analyses of structural abnormalities in ADHD provide significant support implicating the cerebellum, the splenium, basal ganglia, prefrontal cortex, posterior cingulate, and total cerebral volume (Nakao, Radua, Rubia, & Mataix-Cols, 2011; Valera, Faraone, Murray, & Seidman, 2007), and there is some support for structural sex differences. Castellanos and colleagues produced the first papers describing anatomical magnetic resonance imaging (aMRI) to examine brain abnormalities in ADHD girls and to examine ADHD longitudinally in both sexes (Castellanos et al., 2001; Castellanos et al., 2002). In a girls-only sample, total cerebral volume (TCV) was smaller in the ADHD group before correcting for Vocabulary differences on the WISC, but ANCOVA of Vocabulary and TCV did reveal smaller cerebellar vermis and posterior-inferior lobules (Castellanos et al., 2001). Mahone et al., (2011) compared structural differences between ADHD and controls and further explored differences by conducting separate analyses for girls and boys. Boys and girls were compared to sex- and age-matched controls. Furthermore, the authors controlled for hormonal effects by excluding those participants who had reached puberty. Both girls and boys with ADHD evidenced a smaller left supplementary motor cortex (SMC) compared to controls. Boys with ADHD had a smaller left medial prefrontal cortex compared to sex-matched controls, and girls with ADHD had a smaller left lateral premotor cortex. (Mahone et al., 2011). Consistent with neuroimaging research, specific sex differences emerged within response inhibition, such that boys showed abnormalities in areas associated with motor disinhibition and girls showed abnormalities in areas associated with maintenance of response control (Mahone et al., 2011). Furthermore, longitudinal

neuroimaging indicates a pattern of neurodevelopmental delay in many children with ADHD (Shaw et al., 2007).

To put this into context, an eight-year old girl with ADHD may be developmentally behind her same-sex peers, developmentally equivalent to typical eight-year old boys, *and* still developmentally ahead of eight-year old boys with ADHD. While the literature comparing girls to boys with ADHD has focused primarily on gender differences, the dimorphic nature of brain development in girls and boys may mean neuropsychological functioning is better explained in terms of sex differences. In the present study, sex is defined as the structural and functional characteristics of males and females whereas gender is defined as the cultural and behavioral traits of males and females (Torgrimson & Minson, 2005). Therefore, we use the terminology of the literature but signify gender with an * when sex may be more appropriate terminology. As *DSM* criteria for ADHD were established primarily with male-only samples, *DSM* criteria – especially age cutoffs – may provide an insufficient profile of girls with ADHD.

Models of ADHD and Executive Function

Researchers have made appreciable gains toward understanding the gender-specific social, academic, and behavioral profiles of girls with ADHD. However, our understanding of the role neuropsychological deficits play in female ADHD impairment remains conspicuously inadequate. Recent findings suggest 68-78% of children diagnosed with ADHD have impairment in neuropsychological functioning, specifically across executive function (EF) domains (Barkley, 2014), where EF deficits account for 10% of the variance of ADHD symptoms (Willcutt, 2014). Executive function is broadly explained as separate but equal higher order processes related to goal attainment, including (1) response inhibition, (2) working memory, (3) set-shifting/task-shifting, (4) interference control, and planning/organization (Diamond, 2013; Erik G Willcutt,

Alysa E Doyle, Joel T Nigg, Stephen V Faraone, & Bruce F Pennington, 2005; Zelazo & Müller, 2002). In reflection, the *DSM-5* now classifies ADHD as a neurodevelopmental disorder (American Psychiatric Association, 2013), and a number of ADHD models have emerged or been modified to account for high correlations between EF domains and ADHD symptomatology (Rapport, Kofler, Alderson, & Raiker, 2008; Sergeant & van der Meere, 1990). Nevertheless, these models make no attempt to explain sex differences. For example, Barkley (2014) conceptualizes executive functions in a hierarchical developmental framework based largely on the work of (Vygotsky, 1978); however, sexual dimorphism in brain development remains absent from the model.

Existing Literature on ADHD and Executive Function

Given the paucity of research on sex-specific differences in executive function profiles of ADHD, contemporary models should explicate differences and provide key targets for researchers, clinicians, and educators working with girls with ADHD. If models fail to provide predictions for the executive function profile of girls with ADHD, it is important to aggregate the current findings to build more inclusive *and* comprehensive models. To our knowledge, two meta-analytic reviews have compared gender* differences on measures of EF (Gershon, 2002; Hasson & Fine, 2012). Gershon (2002) examined gender* differences on three tasks of EF – the Continuous Performance Task (CPT), the Stroop, and the Matching Familiar Figures Test (MFFT). For the CPT, Gershon (2002) analyzed four studies of Omission Errors (d = -.26) and six studies of Commission Errors (d = .10). For the Stroop, Gershon (2002) calculated the effect sizes for two studies each of Words (d = .07), Colors (d = -.12), and Interference (d = -.19). Five studies yielded two variables for the MFFT – Errors (d = -.15) and Latency (d = -.15). For variables across all tasks, the effect sizes were small, and there were no significant between-

group differences. However, interpretations of these effect sizes are limited given the paucity of available studies. Hasson and Fine (2012) extended Gershon's work on the CPT with the inclusion of five studies of omission and eight studies of commission errors. Omission errors decreased in effect size (d = -.09), while commission errors increased (d = .31). These shifts in effect size between the meta-analyses are not easily clarified as the individual studies used and their calculated effect sizes are not reported (Gershon, 2002). Moreover, there appears to be some inconsistency for the articles included from one meta-analysis to the next. Only two studies from the Hasson and Fine meta-analysis included omission errors and were conducted before 2002; however, we know Gershon calculated results from four studies. Similarly, although Hasson and Fine (2012) included five studies of commission errors conducted before 2002, Gershon (2002) included six. Without further clarity about the studies included in the Gershon meta-analysis, an important question arises about what meaningful changes can be expected with the increase of available studies. Meta-analysis of EF of combined sex ADHD samples may provide some foundation.

Meta-analytic work from Willcutt and colleagues (2005) compared the effect sizes of between-group differences for ADHD and typically developing peers in combined-sex samples on a range of executive functions. The authors proposed a set of criteria which needed to be met in order for executive functions to be considered an integral part of the ADHD profile. These were: (1) consistent executive function weaknesses, even after controlling for confounding variables; (2) the executive function deficits must contribute to a substantial portion of the variance in ADHD symptoms; (3) executive function weaknesses should be present in the majority of ADHD cases; and (4) executive function weaknesses should have common etiology and should be cohertiable with ADHD (Erik G Willcutt et al., 2005). The meta-analysis revealed

a mean weighted effect size of .54 (95% CI = .51-.57) for all measures of executive functions; between-group differences were most consistent for response inhibition (i.e., stop-signal reaction time; *d* =.61), attentional control [cognitive performance tasks (CPT)]; *d* =.51), spatial working memory (*d* =.63), and planning (Tower of Hanoi and Porteus Mazes; *d* =.69 and .58). Shifting (Wisconsin Card Sorting Task; *d* = .46) and Interference (Stroop; *d* = .35) had the lowest mean effect sizes. Subtype differences were negligible (*d* = .09 ±.10); although, the authors proposed possible gender by subtype differences (Nigg, Blaskey, Huang-Pollock, & Rappley, 2002).

Funding Concerns for Current ADHD Research

Despite the paucity of research on sex differences within EF profiles, recent amendments from the National Institute of Health (NIH) require inclusion of both men and women in research. NIH Public Policy 4.1.15.8 (National Institute of Health, 2001) states the following:

> the policy requires that women and members of minority groups and their subpopulations be included in NIH-conducted or supported clinical research, unless a clear and compelling rationale and justification establishes to the satisfaction of the NIH IC Director that inclusion is inappropriate with respect to the health of the subjects or the purpose of the research.

Historically, girls with ADHD are excluded from samples due to the ambiguity surrounding the profile of girls with ADHD (see for review, Hinshaw, Carte, Sami, Treuting, & Zupan, 2002). However, when one considers the reduced ratio of girls to boys with ADHD over time and policy changes from funding agencies, this strategy becomes less feasible. Research on sex differences and EF profiles is needed for researchers to make well-informed decisions when including girls into their research programs. Understanding the EF profile of girls, for example, may inform the development of impairment specific interventions for boys *and* girls with ADHD. If sex-specific EF impairments are noted, for example, a “one sex fits all” intervention may not address sex-specific EF impairments.

Study Aims

The present investigation attempts to (1) synthesize, through a meta-analytic approach, existing literature on sex differences and similarities of EF deficits in children with ADHD and (2) to provide researchers with future directions for examining specific EF deficits in girls with ADHD compared to their same-age female peers.

Chapter 2

Methods

Literature Search and Inclusion Criteria

A search of the literature was conducted according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Moher et al., 2015). A team of seven research assistants conducted a search of the existing literature across the following databases: PsycINFO, PsycARTICLES, Psychiatryonline, PubMed, Google Scholar, ERIC, Sage Journals, and Web of Science. Each research assistant was instructed to utilize four of the eight databases and was given one of the following search term categories: (1) ADHD or atten*; (2) girls or sex differences; and (3) neuro*, cog*, working memory, or executive function*. Search term categories were linked to the modifier AND (e.g.: ADHD AND Girls AND Neuro* or Atten* AND Sex differences AND Cog*). All search term categories were then cross-referenced. An asterisk placed at the end of a word root indicated to database search engines to look for any byproduct of that root (e.g. Cog*: Cognition). Articles were identified through the following inclusion criteria: (1) written in English, (2) peer-reviewed, (3) included a sample of children between the ages of 4 and 18 with ADHD, (4) included both boys and girls in the sample, and 5) published prior to August 2017. Dissertations, theses, and unpublished manuscripts were not considered for the present study. Research assistants met weekly with the principal investigator to provide updates on their progress and to receive ongoing clarification of inclusion criteria. Only the principal investigator had access to all team members' search product. Team members then conducted backward and forward searches of identified articles. A backward search of an article examined reference lists of the selected article, while a forward search used databases (PsycINFO and Google Scholar) to trace articles citing the selected article.

The citations of identified articles were downloaded to EndNote software. Following the initial identification phase, duplicates were located and removed using EndNote's "Find Duplicates" function.

Identification of Studies

Following removal of duplicates, the collected sample studies were combined across research members. Each member screened the full-text of all articles to verify the inclusion of (1) measures of executive function and (2) sufficient data to calculate effect sizes for between-group executive functioning performance. For the purpose of the current meta-analysis, we adhered to the broad definition of executive function as highlighted by Willcutt et al. (2005), which includes response inhibition, attentional control, set-shifting, planning/organization, phonological working memory, and visuospatial working memory. The research team reviewed articles through EndNote citation management software and categorized studies in Microsoft Excel by presence of executive function variables and availability of data to calculate effect sizes. The principal investigator then reviewed the work product of all team members and performed inter-rater reliability calculations. Team members continued to meet weekly to review discrepancies between raters.

In the final process prior to quantitative synthesis of the data, research assistants used Microsoft Excel worksheets to assign study variables and tasks according to their respective executive function domains, as established by Wilcutt and colleagues (2005). To maximize study inclusion, the current meta-analysis diverged from Wilcutt, et al. (2005) by including the following tasks: Matching Familiar Figures error scores (Attentional Control domain), Gordon Diagnostic System commission errors (Inhibition domain), Choice-Delay, delay aversion standard score (Inhibition domain), Stroop, Condition 4 interference T-scores (Set-shifting

domain), Letter-Number Sequencing (Phonological Working Memory domain). The research assistants and principal investigator then targeted a variable from each study which maximized the number of studies available for quantitative analysis in each executive function domain. Longitudinal studies and multiple studies produced from a single data set underwent a decision process whereby the team selected the sample providing maximal quantitative data and matching inclusion criteria. Examples included selecting a sample with the most useable task variables, as well as selecting the sample whose mean age both fit within the 4-18 age criterion and was most consistent with the mean age of other selected studies. As meta-analyses are an iterative process and transparency is critical to replicability, the research team recorded the number of studies included following each step in the data collection process (see Figure 1).

Dependent Variables

Based on the work of Wilcutt et al. (2005), the following dependent variables were included in the present study:

Response inhibition and inhibitory control. Inhibition is often discussed in two forms – attention and action. Attentional control, often referred to as inhibitory control, requires the individual to block-out stimuli irrelevant to attending to one's goal (Diamond, 2013). For example, conversations in public require you to focus on the person with whom you are speaking and exclude sounds from other conversations and atmospheric noises. Inhibition of action, or response inhibition, requires an individual to stop irrelevant and automatic behavioral responses (prepotent responses) irrelevant to one's goal (Diamond, 2013). For example, a young child must inhibit a prepotent response to hit a classmate who takes his toy if he wants to achieve his goal of earning a gold star for the day. Despite the seemingly distinct nature of these two inhibitory components, there appear to be no gross differences (Diamond, 2013). However, within the

literature examining executive dysfunctions, attentional control is often defined as vigilance toward target stimuli, and it is measured as errors of omission on the Continuous Performance Tsk (CPT) (Willcutt et al., 2005). Inhibition of action (i.e., response inhibition) is measured on CPT commission errors, as well as on Stop-Signal Reaction Time (Willcutt et al., 2005).

Set-shifting. Cognitive flexibility requires the ability to shift both between tasks (task-switching) and within tasks (set-shifting; (Diamond, 2013). For example, a child may need to stop watching T.V. and begin getting ready for bed. This is an example of switching between tasks. However, if the child's bedtime is 7:30 p.m. on school nights but 9:00 p.m. on weekends, the child must be able to switch rules flexibly within the task of bedtime. Wilcutt et al. (2005) categorized both the WCST perseverative errors and the Trailmaking Test Part B as tests of cognitive flexibility. The WCST requires participants to use a set of principles for sorting playing cards. The sorting principle changes frequently and without alerting participants. However, participants are alerted to errors they make in sorting; therefore, they much test the remaining sorting principles to obtain correct responses. Participants who have difficulty attending to the error alert and shifting to other possible sorting principles demonstrate weaknesses in cognitive flexibility. Part B of the Trailmaking Test also requires participants to shift back-and-forth between a set of two rules for making a trail (i.e., shifting from letters and numbers).

Planning/organization. Planning is the identification of a goal and breaking the goal down further into a series of strategic and efficient steps to achieve the goal (Best, Miller, & Jones, 2009; Delis, Kaplan, & Kramer, 2001; Denckla, 1994; Georgiou & Das, 2016). Various iterations of the Tower of London task are most commonly attributed to measurement of planning within the executive function umbrella (Borys, Spitz, & Dorans, 1982; Culbertson &

Zillmer, 1998; Delis et al., 2001; Kempton et al., 1999; Shallice, 1982). Tower tasks require participants to use spatial planning to determine the most appropriate and time-efficient moves needed to achieve the goal configuration for stacks of various sized blocks. Wilcutt et al. (2005) also included Porteus Mazes (Krikorian & Bartok, 1998) and the Rey-Osterreith Complex Figure Test (Stern et al., 1994). Within a child's everyday life, planning is involved in many multistep goals, such as getting dressed. A child can get dressed most efficiently if s/he takes out all of the clothes needed for the day, rather than taking each out individually. Inappropriate sequencing of steps might involve the child putting on her shoes before her pants and getting her foot stuck in the leg of her pants.

Working memory. Working memory is the process of holding information in one's awareness and manipulating it in some way to produce a response. It is further theorized to contain two distinct subsidiary systems – the visuospatial sketchpad and the phonological loop (Baddeley & Hitch, 1974). The visuospatial sketchpad processes visual stimuli, while the phonological loop not only processes verbal information but also visual information which a person can tag verbally. An example of this is processing a picture of a chair in the phonological loop as chair, rather than processing the physical characteristics of the chair. In their meta-analysis comparing working memory in children with ADHD to typically developing children, Kasper, Alderson, and Hudec (Kasper, Alderson, & Hudec, 2012) established that the inclusion of girls decreases the between-group differences in working impairments. However, it is possible given the literature on differing neurodevelopmental trajectories for girls and boys, this finding will be moderated by age. The most common phonological working memory tasks are Digits Backwards and Sentence Span. Wilcutt et al. (2005) also identified self-ordered pointing and CANTAB Spatial Working Memory as spatial working memory tasks. For a child, everyday

working memory demands are most commonly found in following directions. Children who have difficulty following a particular sequence of specified steps may have deficits in working memory.

Included Studies

The initial search yielded 148,059 studies through search of PsycINFO, PsycARTICLES, Psychiatryonline, PubMed, Google Scholar, ERIC, Sage Journals, and Web of Science databases. Of the 148,059; 203 records were identified, and 183 studies remained after duplicates were removed (see Figure 1). The abstracts of the 183 studies were evaluated for fit; 135 studies did not meet criteria (i.e., written in a language other than English; did not contain children ages 4-18 in the sample; did not contain girls in the sample). After full review of the remaining 48 articles, an additional 23 articles were removed. Thirteen articles analyzed the same or an overlapping sample of participants (Biederman et al., 2002; Biederman, et al., 2004; Jacobson, et al., 2015; Martinelli, Mostofsky, Stewart, & Rosch, 2016; O'Brien, Dowell, Mostofsky, Denckla, & Mahone, 2010; Oie, et al., 2016; Rosch, Dirlikov, & Mosofsky, 2015; Seidman, et al., 2005; Seymour, Mostofsky, & Rosch, 2016; Skogli, 2013; Skogli, Teicher, Andersen, Hovik, & Oie, 2014; Wodka et al., 2008a; Wodka et al., 2008b). To prevent double dipping (i.e., counting the same participants more than once), coders evaluated the articles based on the number of participants and the number and kind of analyses undertaken. If articles provided findings for different EF variables, both articles were included in analyses. Otherwise, articles were selected to maximize findings (e.g., article with most EF variables, article with largest sample of participants). For example, from the Johns Hopkins sample (Jacobson et al., 2015; Martinelli et al., 2016; O'Brien et al., 2010; Rosch et al., 2015; Seymour et al., 2016), O'Brien et al. (2010) provided the most EF variables (i.e., planning, shifting, inhibition, and working memory) despite

having the fewest number of participants. Seymour et al. (2016) provided a more traditional measure of inhibition (i.e., Go/NoGo commission errors) and provided omission errors; therefore, it was used to calculate effect sizes for inhibition and attentional control. Two studies mentioned conducting sex analyses but did not report statistics (Houghton et al., 1999; Hutchison, Feder, Abar, & Winsler, 2016). Nine articles failed to report sex-specific ADHD means and standard deviations (*SD*s) or analyses of interactions by sex (Bezidijan, Baker, Lozano, & Raine, 2009; Booth et al., 2005; Rohrer-Baumgartner et al., 2014; Rubia, Smith, & Taylor, 2007; Skogan et al., 2014; Sonuga-Barke, Bitsakou, & Thompson, 2010; Uebal et al., 2010; van Ewijk et al., 2014; Willfors et al., 2014). Two articles were removed because they provided only beta weights for interactions of sex and diagnostic group on EF tasks (Hartung, Millich, Lynman, & Martin, 2002; Sarkis, Sarkis, Marshall, & Archer, 2005). Four studies examined ADHD symptoms as a continuous variable but did not utilize an ADHD diagnostic group (Gray, Rogers, Martinussen, & Tannock, 2015; Kallitsogolou, 2013; Michel, Molitor, & Schneider, 2016; and Piek et al., 2004). One study examined a composite EF measure (McQuade, Breaux, Miller, & Mathias, 2016). The remaining 22 studies yielded 45 variables (Attentional control, n = 10; Response Inhibition, n = 12; Set-Shifting, n =13; Planning, n = 6; Phonological Working Memory, n = 4). Spatial working memory was available for only 2 studies (O'Brien et al., 2010; Rucklidge, 2006), and, therefore, could not be examined.

Chapter 3

Analyses

Calculating Effect Sizes

The following meta-analysis used a random-effects model because we assumed the true effect of each study varied according to significant differences in sample characteristics (e.g. mean age) and measurement of executive function variables. Dichotomous independent groups, girls with ADHD and boys with ADHD, were compared on continuous dependent measures of executive function. For data including dichotomous independent variables and continuous dependent variables, mean differences were calculated (Cooper, Hedges, & Valentine, 2009). As executive function domains are assessed with a variety of experimental tasks and clinical measures, mean differences were standardized, *d* statistic, as opposed to using raw means, *D* statistic (Cooper et al., 2009). There are three primary formulas for the *d* statistic – Cohen's *d*, Hedge's *g*, and Glass's *delta*. Cohen's *d* is generally used for studies with greater than 20 participants in each sample (Cooper et al., 2009). Means, standard deviations (*SD*s), and sample sizes for each study and variable were entered into Comprehensive Meta-Analysis (v3), and Cohen's *d* and 95% confidence intervals were computed. One study did not provide means and *SD* (Loo et al.,); however, the researchers did provide a sex × diagnostic group interaction effect using Cohen's *F*. As such, Cohen's *d* was then calculated using a formula from Cohen (1988).

Assumptions of Meta-Analyses

The ability to compare standardized effect sizes is based on the assumption that confidence intervals of study effects differ significantly only as a product of chance. We are able to test this assumption using Cochran's x^2 test, or the *Q*-test (Higgins & Thompson, 2002). The *Q*-statistic evaluates the presence of heterogeneity existing beyond that which we would expect

with chance and suggests measurement differences and/or study design are an additional factor for studies with non-overlapping confidence intervals (*Q* STATISTIC). The *p*-value of the combined effect size, although often cited, provides insufficient power of effect size to meta-analyses of few studies and excessive power to meta-analyses combining a large number of studies (Higgins & Thompson, 2002). Still, factors influencing *Q* include (1) the presence of heterogeneity, (2) the number of studies, and (3) the weights given to each study (Hardy & Thompson, 1998). Should study effects differ beyond chance, we are then able to test, through meta-regression, moderators which might account for these differences. Meta-analyses are additionally subject to bias toward a meaningful effect due to bias of published research toward significant findings (Rosenberg, 2005). Therefore, a meta-analysis cannot satisfy the assumption of combining all studies examining the variable to interest without accounting for publication bias. The calculation of Fail-Safe Numbers (Fail-Safe *N*) allows us to define the number of unpublished, nonsignificant studies needed to render the meta-analysis results insignificant (Rosenberg, 2005). Broadly speaking, fewer studies are needed to detect large effect sizes because the probability of detecting significant results within a given sample is higher. Conversely, it may be necessary to draw many samples from the population to detect small effect sizes (i.e., results which are present in only a small amount of the population). The current meta-analysis calculated Rosenthal's Fail-Safe *N* (Rosenthal, 1979; Begg, 1994).

Meta-Regression

Significant variability across studies for a chosen effect size can often be explained due to differences in samples and/or in methodology. Meta-regression analyses (Thompson & Higgins, 2002), the degree to which specific variables of sample and methodological diversity moderate

the strength of the effect, is a quantitative method for assessing the impact of methodology differences.

Thompson and Higgins (2002) follow findings from Hardy and Thompson (1998) indicating tests of heterogeneity have relatively low power and suggest meta-regression analyses are appropriate even for non-significant results of heterogeneity tests. Prior to meta-regression analyses, studies used for each EF variable were coded for potential covariates. Coding was conducted by two research assistants. The team resolved ambiguous definitions, either of covariate or coding scheme, and re-coded. Interrater reliability coefficients were calculated until the team achieved interrater agreements of at least 80%. The final coding iteration ended at 85% reliability. The lead investigator resolved remaining discrepancies. Broadly, the investigator evaluated whether moderators reflected a more general or a more specific result. For example, if studies did not specify ADHD subtype distribution for the sample, the sample was coded as including a mixed sample.

To protect against false positives, the present study predefined the following covariates. For sample differences, moderators of interest include: age, IQ, and whether participants were drawn from a clinical or community sample. Methodological moderators included: the number of raters used to determine ADHD diagnosis, whether ADHD samples combined ADHD subtypes or examined ADHD-Combined samples only, as well as whether ADHD participants underwent a wash-out period for stimulant medication prior to completion of EF measures. See Appendix 3 for Codebook Manual. As there were few studies in each domain and testing multiple covariates has the potential to identify false positives (Thompson & Higgins, 2002), the current meta-analysis was unable to test the effects of moderator variables. However, moderators are presented in Table 1, 2, 3, 4, and 5 for each domain.

Chapter 4

Results

Analysis of Effect Size Distribution for Attentional Control

Analyses examined nine effect sizes, ranging from -.344 to 1.284 (see Table 1). The overall effect size comparing girls with ADHD to boys with ADHD on omission errors was not significant (d = -0.071, p = 0.417), suggesting no difference in omission errors between sexes. The 95% Confidence Interval ranged from -0.242 to 0.100. A non-significant test of heterogeneity (Q-value = 9.929, p = 0.270) indicated studies did not differ beyond what would be expected by sampling error. As the mean effect size indicated no difference in performance between girls and boys with ADHD, it follows that the Fail-Safe N method determined <0.001 additional studies would increase the p-value above alpha.

Analysis of Effect Size Distribution for Inhibition

Analyses examined 11 effect sizes, ranging from -.571 to 0.869 (see Table 2). The positive effect size comparing girls with ADHD to boys with ADHD on commission errors was small and not significant (d = -0.102[1], p = 0.319), suggesting no inhibitory differences between boys and girls. The 95% Confidence Interval ranged from -0.098 to 0.302. A significant test of heterogeneity (Q-value = 19.880, p = 0.030) indicated studies did not differ beyond what would be expected by sampling error. The Fail-Safe N method was significant for publication bias and determined only eight additional studies would increase the p-value above alpha.

[1] Positive effect sizes indicate girls with ADHD performed better than did their male counterparts. Negative effects indicate poorer performance for girls.

Analysis of Effect Size Distribution for Set-Shifting

Analyses examined 13 effect sizes, ranging from -1.347 to 0.484 (see Table 3). The overall effect size comparing girls with ADHD to boys with ADHD on set-shifting was not significant (d = 0.157, p = 0.101), suggesting no difference in set-shifting performance between sexes. The 95% Confidence Interval ranged from -0.031 to 0.344. A significant test of heterogeneity (Q-value = 25.732, p = 0.012) indicated studies differed beyond what would be expected by sampling error. Consistent with a null effect for mean differences between girls and boys with ADHD on set-shifting, Fail-Safe N determined <0.001 additional studies would increase the p-value above alpha.

Analysis of Effect Size Distribution for Planning/Organization

Analyses examined six effect sizes across six samples, ranging from -0.684 to 0.404 (see Table 4). The overall effect size comparing girls with ADHD to boys with ADHD on planning/organization abilities was small and not significant (d = 0.009, p = 0.944). Four of six studies revealed higher mean planning performance for girls than boys with the disorder (DeShazo, 2000; O'Brien et al., 2010; Skogli et al., 2013; Siedman et al., 2005); however, only Lockwood et al. (2001) found significant sex differences and showed boys outperformed girls. The 95% Confidence Interval ranged from -0.161 to 0.205. A non-significant test of heterogeneity (Q-value = 8.214, p = 0.145) indicated studies did not differ beyond what would be expected by sampling error. The Fail-Safe N method determined <0.001 additional studies would increase the p-value above alpha.

Analysis of Effect Size Distribution for Phonological Working Memory

Analyses examined four effect sizes, ranging from -0.028 to 0.256 (see Table 5). The overall effect size comparing girls with ADHD to boys with ADHD on working memory

performance was small and not significant (d = 0.146, p = 0.210), suggesting no difference in working memory between sexes. The 95% Confidence Interval ranged from -0.082 to 0.375. A non-significant test of heterogeneity (Q-value = 0.602, p = 0.896) indicated studies did not differ beyond what would be expected by sampling error. The Fail-Safe N method determined only one additional study would increase the p-value above alpha.

Chapter 5

Discussion

Summary

The present meta-analysis attempted to synthesize existing literature on sex differences and similarities of EF deficits in children with ADHD. Previous research (Gaub & Carlson, 1997; Gershon, 2002) found girls and boys do differ in symptomatology; however, our analyses indicated no differences on measures of executive functioning. Existing literature on EF sex differences remains sparse, and only 22 studies were available for analyses. Nevertheless, the current data represent an extensive international sample with studies from Canada (Poissant, Rapin, Chenail, & Mendrek, 2016; Rucklidge & Tannock, 2002), Egypt (Roufael et al., 2012), Finland (Loo et al., 2007), Germany (Gunther, Knospe, Herpert-Dahlmamn, & Konrad, 2015), Iran (Heydari & Farahani, 2016), New Zealand (Rucklidge, 2006), Norway (Skogli et al., 2013), Sweden (Sjowall, Roth, Lindqvist, & Thorell, 2013), Taiwan (Yang, Jong, Chung, & Chen, 2014), and the U.S. (Ackerman, Roscoe, Dykman, & Oglesby, 1983; Boseck, Davis, Cassady, Finch, & Gieder, 2016; Breen, 1989; DeShazo, 2000; Horn, Wagner, & Ialongo, 1989; Lockwood, Marcotte, & Stern, 2001; Newcorn et al., 2001; O'Brien et al., 2010; Seidman, et al., 2005; Seymour et al., 2016; Wodka et al., 2008a).

Attentional control. Consistent with previous meta-analyses (Gershon, 2002; Hasson & Fine, 2012), there were no significant sex differences for attentional control performance. As there was not a significant difference, moderator analyses could not be run. However, there were notable qualitative differences between the studies. Seven studies tested omission errors using the CPT (Loo et al., 2007; Newcorn et al., 2001; Poissant et al., 2016; Rucklidge, 2006; Siedman et al., 2005; Horn et al., 1989; and Yang et al., 2004). Two studies used the Simple Go/No-Go

(Seymour et al., 2016; Wodka et al., 2008), and one used the MFFT (Ackerman, 1983). Eight studies showed girls performing slightly worse than boys with ADHD. Only Poissant et al. (*N* = 23; 2016) found a statistically significant difference between the sexes, indicating girls outperformed boys with ADHD.

Inhibition. Inconsistent with findings from previous meta-analyses that examined CPT commission errors only (Gershon, 2002; Hasson & Fine, 2012), boys and girls did not differ on inhibitory control. Eight of the 11 studies examined commission errors on the CPT. However, two studies were conducted using a simple Go/No-Go paradigm (Seymour, et al., 2016; Wodka et al., 2007), one study examined a Choice-Delay task (Sjowall, 2013), one study measured reaction time on the Stop-Signal task (Rucklidge & Tannock, 2002), and one study was conducted using the Gordon Diagnostic System (Breen, 1989). Furthermore, Newcorn et al. (2001) took a more nuanced approach to calculating commission errors in that they defined commissions as, "two types of commission errors: those in which A was not followed by X, with short RT, and A-only errors with long RT," (p. 139; Newcorn et al., 2001). For the Choice Delay Task (Sonuga-Barke, Taylor, Sembi, & Smith, 1992), children could choose an immediate, smaller reward or wait for a larger reward. The addition of a reward may measure a more emotional form of inhibition and therefore tap what is considered a "hot" EF (Zelazo & Muller, 2002). "Hot" EF's refer to affective decision making and includes tasks leading to meaningful goals and rewards (Zelazo, Qu, & Muller, 2005). Notably, this task evidenced one of the lowest effect sizes. Sample characteristics also varied considerably. Poissant et al. (2016) conducted analyses on a sample of only 23 children with ADHD. Siedman et al. (2005) had the largest sample at 204 children with ADHD. Samples from Loo et al. ([age range = 16 — 18] 2007), Rucklidge ([girls *M* age = 14.94, *SD* = 1.22; and boys *M* age = 14.78, *SD* = 0.99] 2006), and

Rucklidge and Tannock ([girls *M* age = 14.68, *SD* = 1.51; and boys *M* age = 14.80, *SD* = 1.22] 2001) were older, on average, than were other samples.

Set-shifting. Gershon (2002) previously examined set-shifting performance across just two studies using the Stroop Interference variable; the effect size was small and not significant (*d* = -0.19, *p* = 0.16). Our meta-analysis also failed to find sex differences, and current findings are notable for extending the number of tasks and samples included. Six of 13 studies provided error scores: WCST, perseverative errors (Roufael et al., 2012; Rucklidge, 2006; Siedman et al., 2005), Stroop, Condition 4 interference *T*-score (Rucklidge & Tannock, 2002), and TMT, Test B errors (DeShazo et al., 2000), and Navon-like task, set-shifting errors (Sjowall et al., 2013). Six studies provided variables where higher means indicated better performance: TMT, Test B standard score (Lockwood & Stern, 2001), TMT, Condition 4 standard score (O'Brien et al., 2010; Skogli et al., 2013), WCST standard score (Heydari & Farahani, 2016), and a visual set-shifting task (de Sonneville, 2000) measuring efficiency (Gunther et al., 2015). Three studies found girls performed worse on set-shifting than did boys (DeShazo et al., 2000; O'Brien et al., 2010; and Siedman et al., 2005). Studies indicating girls performed more poorly did not share task or variable similarities.

Planning/organization. This meta-analysis was the first, to our knowledge, to examine ADHD sex differences in planning/organization. Across six studies, there were no significant differences on planning performance. However, results should be interpreted with caution due to the small number of studies included. Furthermore, the present analyses encompass a wide variety of tasks assessing planning. Specifically, three tests examined Tower of London (ToL) performance. One study (DeShazo et al., 2000) examined Tower of Hanoi (ToH) performance, which is a computerized adaptation of the traditional ToL task (Borys et al., 1982; Leon-Carrion,

1999). Two studies reported standard scores for Copy Organization on the Rey-Ostererich Complex Figures Task (Lockwood, Marcotte, & Stern, 2001; Siedman, et al. 2005). All but one study indicated girls outperform boys on planning tasks (Lockwood et al., 2001); this study differed from others in that it included only children with ADHD-C, rather than including multiple subtypes.

Phonological Working memory. A previous meta-analysis found that inclusion of females moderated the effect of working memory performance between children with and without ADHD, such that differences were minimized when females were included. Several research groups have cited this meta-analysis as a rationale for excluding females when studying working memory. The current meta-analysis did not find similar results; specifically, girls and boys did not differ on phonological working memory tasks. However, our meta-analysis measured two working memory tasks across only four studies: letter-number sequencing (Boseck et al., 2015; Skogli et al., 2013) and digits span backwards (Lockwood et al., 2001; O'Brien et al., 2010). We were also unable to assess visuospatial working memory and/or central executive functioning. Therefore, comparisons between the meta-analyses may not be comparable. Furthermore, Kasper et al. (2012) also found that inclusion of less than 10 trials per set size decreased group differences between ADHD and typically developing children. All of the studies included in our meta-analysis had only two trials per set size; therefore, it is possible sex differences, like group differences, emerge across a greater number of trials.

Conclusions

Prior reviews of sex differences in ADHD have emphasized ways females may be uniquely affected by ADHD given that the disorder is primarily diagnosed in males (Mahone & Wodka, 2008; Nussbaum, 2012). Specifically, they cite the work of Eme (1992) which proposes

that the sex least affected by a given disorder will manifest greater impairments. However, this and previous meta-analyses have failed to find clear support for this theory (Gaub & Carlson, 1997; Gershon, 2002; Hasson & Fine, 2012). Instead, it appears girls and boys do not differ substantially in their degree of EF impairment for attentional control, set-shifting, planning/organization, or phonological working memory on the tasks included; and boys perform worse on measures of inhibition. One alternative hypothesis may be that while early maturation (Shaw et al., 2007) may protect girls from early presentation, the effects of puberty are likely to lead to more impairment, post-adolescence, due to increased dopamine receptors (Keltner &Taylor, 2002; Mahone & Wodka, 2008). Alternatively, the development of comorbid internalizing and externalizing disorders may exacerbate EF profiles.

Current research into executive functions and ADHD may be inadequate. Milich, Hartung, Martin, and Haigler (1994) noted a significant problem of circular logic expressly in regard to research on ADHD and impulsivity although it may be applicable to neuropsychological constructs and application to clinical populations broadly. Commonly, neuropsychological tasks are first observed to distinguish between healthy and clinical populations. Research is then designed using these neuropsychological tasks, and the tasks then validate the neuropsychological construct. However, there is no universally supported model either for ADHD nor for executive functions. What is more, no model of ADHD conceptualizes the disparity between the sexes, and research demonstrating the sensitivities of neuropsychological tasks to ADHD impairments are far and away conducted on predominately or exclusively male samples. Taken together, the issue of circular logic is particularly salient for the present meta-analysis.

Limitations

Despite the power of meta-analyses to aggregate effects across samples and differing methodological approaches, they are not without their limitations. Meta-analyses can only make generalizations about effects from existing research. Therefore, the quality and the accuracy of meta-analytic synthesis is dependent upon the quality and breadth of the current literature.

The dimorphic trajectories of males and females (Shaw et al., 2006) suggests sex-specific comparisons between control and ADHD groups will provide greater insight into the comparatively small number of girls to boys diagnosed with ADHD than will sex comparisons within an ADHD group. However, the paucity of comparisons between girls with and without ADHD within extant literature precluded the possibility of meta-analytic synthesis. What is more, although analyses of potential sex differences were possible, quantitative examination of potential moderators was not.

Boundaries in scope increase the feasibility of meta-analyses; however, they also serve as limitations. The scope of the present meta-analysis was confined to research published before October, 2017. Therefore, the current findings are limited by publication bias and the availability of studies on or before the stop date of the literature search. The present meta-analysis also defined EF domains and their corresponding tasks according to the work of Willcutt and colleagues (2005). It is possible that consolidation, expansion, or reconceptualization of EF domains; the use of different neuropsychological tasks; and/or the use of different task variables would alter the magnitude of effects between sexes.

Finally, the present meta-analysis aimed to obtain the greatest number of studies per EF domain. As such, inclusion was open to those studies whose diagnostic groups were established with early editions of the DSM. Of the studies included, DSM editions ranged from DSM-III to

DSM-5. From DSM-IV-TR to DSM-5, the age of symptom onset from 7 to 12 and the conceptualization of ADHD as a neurodevelopmental disorder are notable alterations to ADHD criteria. Consequently, research with earlier editions of the DSM may have under identified individuals with presentations of symptoms in preadolescence. A variety of tasks corresponding to each domain was also permitted. However, task was not examined as a potential moderator. Consequently, the magnitude of effects may be impacted by task.

Future Directions

The current meta-analysis does not support significant EF differences between girls and boys diagnosed with ADHD. Neurodevelopmental research finds a sexual dimorphism in brain development. It was, consequently, of interest to the current study to examine potential age-related changes. However, given that relatively few studies were included, we were unable to assess age as a potential moderator. Furthermore, studies included girls in much fewer numbers than boys, and our meta-analysis parsed out EF into five EF domains modeled after the work of Willcutt and colleagues (2005). Therefore, more work is needed to verify that EF differences between girls and boys with ADHD truly do not exist. Future research should include girls in adequate numbers to appropriately test the relationship of sex by executive function outcome, as well as sex-specific analyses between typically and atypically developing children. It is possible differences will emerge within a single sex, such that girls with ADHD manifest greater EF impairment compared to their typically developing peers. Furthermore, it is possible girls with ADHD will exhibit EF impairments compared to typically-developing girls on tasks (e.g., set-shifting, phonological working memory) not present in broad comparisons of children with ADHD to their typically developing peers. Seminal factor analytic work demonstrated a highly correlated but distinct three-factor structure for traditional neuropsychological measures ([i.e.,

"Shifting," "Updating," and "Inhibition"] Miyake, Friedman, Emerson, Witzki, & Howerter, 2000). As such, current EF domain distinctions may be inappropriate. It is, thus, integral to our understanding of the neuropsychological profiles of girls and boys with ADHD for future work to examine domain-general EF. Similarly, developmental literature suggests girls outperform boys on measures of verbal fluency (Brocki & Bohlin, 2004; Reader, Harris, Schuerholz, & Denckla, 1994). As such, future research should examine differences between girls and boys with ADHD on verbally-mediated tasks. Reconceptualization of executive functions in these ways and availability of data to examine age as a moderator may yield sex differences.

Despite growing interest into female ADHD research, the current body of literature remains scant. Future research should compare girls with ADHD to their typically developing peers and compare these results to the magnitude of effect sizes for sex within and between diagnostic groups. Particular attention to traditional objective measures of attention and inhibition (i.e., CPT, TOVA, SSRT) may be warranted given their clinical popularity for differential diagnosis (Epstein, Erkanil, Conners, Klaric, Costello, & Angold, 2002). If girls with ADHD do not manifest significant impairment on these tasks relative to their peers and/or exhibit more notable impairments on other EF tasks, ADHD may go under-diagnosed in girls with the disorder. Researchers should test the sensitivity of both commonly and uncommonly used neuropsychological tasks to identify group differences. However, important research is also needed to investigate how clinicians prioritize data when making an ADHD diagnosis. If clinicians are not making decisions based on the known profiles of ADHD and/or present a sex bias in their diagnostic methodology, this would have serious implications to ADHD prevalence in girls. Leading theorists, researchers, and clinicians must also investigate and define deficits and characteristics of ADHD most predictive of functional impairment. For example, studies

may wish to compare objective EF measures to academic and intellectual performance. If objective performance measures are weakly correlated with academic and relational impairment, clinicians need instruction and availability of more appropriate diagnostic tools.

Furthermore, although girls continue to be diagnosed in greater numbers, there remains a significant discrepancy between the number of affected males and females. Thusly, it is important for models of ADHD to provide some hypotheses about the etiology of ADHD that speak to the sex discrepancy in diagnosis.

References

Abikoff, H. (1991). Cognitive training in ADHD children: Less to it than meets the eye. *Journal of learning Disabilities, 24*(4), 205-209.

Ackerman, D. (1983). Sex and group differences in reading and attention disordered children with and without hyperkinesis.

American Psychiatric Association. (2013). Neurodevelopmental Disorders *Diagnostic and Statistical Manual of Mental Disorders* (5th ed., pp. 59-66). Washington, D.C.

Baddeley, A. D., & Hitch, G. (1974). Working memory. *Psychology of learning and motivation, 8*, 47-89.

Barkley, R. A. (2014). Executive Functioning and Self-Regulation Viewed as an Extended Phenotype: Implications of the Theory for ADHD and Its Treatment. In R. A. Barkley (Ed.), *Attention-Deficit Hyperactivity Disorder: A Handbook for Diagnosis and Treatment* (Fourth ed., pp. 405-434). New York: Guilford Press.

Best, J. R., Miller, P. H., & Jones, L. L. (2009). Executive functions after age 5: Changes and correlates. *Developmental review, 29*(3), 180-200.

Borys, S. V., Spitz, H. H., & Dorans, B. A. (1982). Tower of Hanoi performance of retarded young adults and nonretarded children as a function of solution length and goal state. *Journal of Experimental Child Psychology*, *33*(1), 87-110.

Breen, M. J. (1989). Cognitive and behavioral differences in AdHD boys and girls. *J Child Psychol Psychiatry, 30*, 711-716.

Cooper, H. M., Hedges, L. V., & Valentine, J. C. (2009). *The handbook of research synthesis and meta-analysis*. New York: Russell Sage Foundation.

Culbertson, W. C., & Zillmer, E. A. (1998). The Tower of London DX: A standardized approach to assessing executive functioning in children. *Archives of Clinical Neuropsychology, 13*(3), 285-301.

Delis, D. C., Kaplan, E., & Kramer, J. H. (2001). *Delis-Kaplan Executive Function System: Examiners manual*: Psychological Corporation.

Denckla, M. B. (1994). Measurement of executive function. In G. R. Lyon (Ed.), *Frames of reference for the assessment of learning disabilities: New views on measurement issues* (pp. 117-142). Baltimore: Paul H. Brookes Publishing.

DeShazo, T. M. (2000). Executive Functions in Boys and Girls with Attention-Deficit/Hyperactivity Disorder: Academic Implications and Potential Benefits of Stimulant Medication.

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

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental psychology, 40*(2), 177.

Gaub, M., & Carlson, C. (1997). Gender differences in ADHD: A meta-analysis and critical review. *Journal of the American Academy of Child & Adolescent Psychiatry, 36*(8), 1036-1045.

Georgiou, G. K., & Das, J. (2016). What component of executive functions contributes to normal and impaired reading comprehension in young adults? *Research in developmental disabilities, 49*, 118-128.

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

Hardy, R. J., & Thompson, S. G. (1998). Detecting and describing heterogeneity in meta-analysis. *Statistics in medicine, 17*(8), 841-856.

Hasson, R., & Fine, J. G. (2012). Gender differences among children with ADHD on continuous performance tests: A meta-analytic review. *Journal of Attention Disorders, 16*(3), 190-198.

Higgins, J., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in medicine, 21*(11), 1539-1558.

Horn, Wagner, & Ialongo. (1989). Sex differences in school-aged children with pervasive attention deficit hyperactivity disorder.

Huizinga, M., & van der Molen, M. W. (2007). Age-group differences in set-switching and set-maintenance on the Wisconsin Card Sorting Task. *Developmental neuropsychology, 31*(2), 193-215.

Kasper, L. J., Alderson, R. M., & Hudec, K. L. (2012). Moderators of working memory deficits in children with attention-deficit/hyperactivity disorder (ADHD): a meta-analytic review. *Clinical psychology review, 32*(7), 605-617.

Kempton, S., Vance, A., Maruff, P., Luk, E., Costin, J., & Pantelis, C. (1999). Executive function and attention deficit hyperactivity disorder: stimulant medication and better executive function performance in children. *Psychological medicine, 29*(3), 527-538.

Krikorian, R., & Bartok, J. A. (1998). Developmental data for the Porteus maze test. *The Clinical Neuropsychologist, 12*(3), 305-310.

Leon-Carrion, J. (1999). Seville computerized neuropsychological test battery (BNS). *Madrid: TEA.*

Lockwood, K. A., Marcotte, A. C., & Stern, C. (2001). Differentiation of attention-deficit/hyperactivity disorder subtypes: application of a neuropsychological model of attention. *J Clin Exp Neuropsychol, 23*(3), 317-330. doi:10.1076/jcen.23.3.317.1179

Loo, S. K., Humphrey, L. A., Tapio, T., Moilanen, I. K., McGOUGH, J. J., McCRACKEN, J. T., . . . Ebeling, H. (2007). Executive functioning among Finnish adolescents with attention-deficit/hyperactivity disorder. *Journal of the American Academy of Child & Adolescent Psychiatry, 46*(12), 1594-1604.

Mahone, E. M., & Wodka, E. L. (2008). The neurobiological profile of girls with ADHD. *Developmental disabilities research reviews, 14*(4), 276-284.

Milich, R., Hartung, C. M., Martin, C. A., & Haigler, E. D. (1994). Behavioral disinhibition and underlying processes in adolescents with disruptive behavior disorders *Disruptive behavior disorders in childhood* (pp. 109-138): Springer.

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., . . . Stewart, L. A. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic reviews, 4*(1), 1.

National Institute of Health. (2001). *Inclusion of Women and Minorities as Subjects in Clinical Research and Reporting Sex/Gender, Racial, and Ethnic Participation*. (DHHS-4.1.15.8). Retrieved from https://grants.nih.gov/grants/policy/nihgps/html5/section_4/4.1_public_policy_requirements_and_objectives.htm#Inclusion.

Newcorn, J. H., Halperin, J. M., Jensen, P. S., Abikoff, H. B., Arnold, L. E., Cantwell, D. P., . . . Greenhill, L. L. (2001). Symptom profiles in children with ADHD: effects of comorbidity

and gender. *Journal of the American Academy of Child & Adolescent Psychiatry, 40*(2), 137-146.

Nigg, J. T., Blaskey, L. G., Huang-Pollock, C. L., & Rappley, M. D. (2002). Neuropsychological executive functions and DSM-IV ADHD subtypes. *Journal of the American Academy of Child & Adolescent Psychiatry, 41*(1), 59-66.

Nussbaum, N. L. (2012). ADHD and female specific concerns: a review of the literature and clinical implications. *Journal of Attention Disorders, 16*(2), 87-100.

Poissant, H., Rapin, L., Chenail, S., & Mendrek, A. (2016). Forethought in Youth with Attention Deficit/Hyperactivity Disorder: An fMRI Study of Sex-Specific Differences. *Psychiatry Journal, 2016*.

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

Rapport, M. D., Kofler, M. J., Alderson, R. M., & Raiker, J. (2008). Attention-deficit/hyperactivity disorder. In M. Hersen & D. Reitman (Eds.), *Handbook of psychological assessment, case conceptualization and treatment, volume 2: Children and adolescents* (pp. 349-404). New Jersey: Wiley.

Rosenberg, M. S. (2005). The file-drawer problem revisited: a general weighted method for calculating fail-safe numbers in meta-analysis. *Evolution, 59*(2), 464-468.

Rucklidge, J. J. (2006). Gender Differences in Neuropsychological Functioning of New Zealand Adolescents with and without Attention Deficit Hyperactivity Disorder. *International Journal of Disability, Development and Education, 53*, 47-66. doi:10.1080/10349120600577402

Rucklidge, & Tannock. (2001). "Psychiatric, Psychosocial, and Cognitive Functioning of Female Adolescents With ADHD".

Sciutto, M. J., Nolfi, C. J., & Bluhm, C. (2004). Effects of child gender and symptom type on referrals for ADHD by elementary school teachers. *Journal of Emotional and Behavioral Disorders, 12*(4), 247-253.

Seidman. (2005). Impact of Gender and Age on Executive Functioning: Do Girls and Boys With and Without Attention Deficit Hyperactivity Disorder Differ Neuropsychologically in Preteen and Teenage Years?

Sergeant, J., & van der Meere, J. (1990). Convergence of approaches in localizing the hyperactivity deficit. In A. Kazdin & B. B. Lahey (Eds.), *Advances in clinical child psychology* (pp. 207-246). New York: Springer.

Seymour, K. E., Mostofsky, S. H., & Rosch, K. S. (2016). Cognitive load differentially impacts response control in girls and boys with ADHD. *Journal of Abnormal Child Psychology, 44*(1), 141-154.

Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 298*(1089), 199-209.

Shaw, P., Eckstrand, K., Sharp, W., Blumenthal, J., Lerch, J., Greenstein, D., . . . Rapoport, J. (2007). Attention-deficit/hyperactivity disorder is characterized by a delay in cortical maturation. *Proceedings of the National Academy of Sciences, 104*(49), 19649-19654.

Shaw, P., Lerch, J., Greenstein, D., Sharp, W., Clasen, L., Evans, A., . . . Rapoport, J. (2006). Longitudinal mapping of cortical thickness and clinical outcome in children and adolescents with attention-deficit/hyperactivity disorder. *Archives of general psychiatry, 63*(5), 540-549.

Sjöwall, D., Roth, L., Lindqvist, S., & Thorell, L. B. (2013). Multiple deficits in ADHD: executive dysfunction, delay aversion, reaction time variability, and emotional deficits. *Journal of Child Psychology and Psychiatry, 54*(6), 619-627.

Sonuga-Barke, E., & Taylor, E. (1992). The Effect of Delay on Hyperactive and Non-hyperactive Children's Response Times: a Research Note. *Journal of Child Psychology and Psychiatry, 33*(6), 1091-1096.

Stern, R. A., Singer, E. A., Duke, L. M., Singer, N. G., Morey, C. E., Daughtrey, E. W., & Kaplan, E. (1994). The Boston qualitative scoring system for the Rey-Osterrieth complex figure: description and interrater reliability. *The Clinical Neuropsychologist, 8*(3), 309-322.

Thompson, S. G., & Higgins, J. (2002). How should meta-regression analyses be undertaken and interpreted? *Statistics in medicine, 21*(11), 1559-1573.

Torgrimson, B. N., & Minson, C. T. (2005). Sex and gender: what is the difference? , *99*(3), 785-787.

Vygotsky, L. (1978). Interaction between learning and development. *Readings on the development of children, 23*(3), 34-41.

Waschbusch, D. A., & King, S. (2006). Should sex-specific norms be used to assess attention-deficit/hyperactivity disorder or oppositional defiant disorder? *Journal of consulting and clinical psychology, 74*(1), 179.

Willcutt, E. G. (2014). Theories of ADHD. In R. A. Barkley (Ed.), *Attention-deficit hyperactivity disorder: a handbook for diagnosis and treatment* (Fourth ed., pp. 355-404). New York: The Guilford Press.

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

Willcutt, E. G., Doyle, A. E., Nigg, J. T., Faraone, S. V., & Pennington, B. F. (2005). Validity of the Executive Function Theory of Attention-Deficit/Hyperactivity Disorder: A Meta-Analytic Review. *Biological psychiatry*, *57*(11), 1336-1346.

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

Wodka, E. L., Mark Mahone, E., Blankner, J. G., Gidley Larson, J. C., Fotedar, S., Denckla, M. B., & Mostofsky, S. H. (2007). Evidence that response inhibition is a primary deficit in ADHD. *Journal of clinical and experimental neuropsychology*, *29*(4), 345-356.

Yang, P., Jong, Y. J., Chung, L. C., & Chen, C. S. (2004). Gender differences in a clinic-referred sample of Taiwanese attention-deficit/hyperactivity disorder children. *Psychiatry and clinical neurosciences*, *58*(6), 619-623.

Zelazo, P. D., & Müller, U. (2002). Executive function in typical and atypical development. In U. Goswami (Ed.), *Blackwell handbook of childhood cognitive development*. Malden: Blackwell Publishing.

Appendices

Appendix A

Tables

Table 1

Effect Sizes and Moderators for Attentional Control

Study	Year	Girls *N*	Age *M*(*SD*)	FSIQ *M*(*SD*)	Boys *N*	Age *M*(*SD*)	FSIQ *M*(*SD*)	Cohen's (*d*)	Referral Source	Diagnostic Raters	Subtype	Medication Washout
Ackerman et al.	1983	9	8.53 (N/R)	102.60 (N/R)	24	8.40 (N/R)	107.90 (N/R)	-0.093	Community	Single Rater	ADHD-H/I	≥24 hours
Horn et al.	1989	17	8.20 (1.30)	N/R	37	8.10 (1.50)	N/R	-0.344	Clinical	Multirater	ADHD-H/I	≥24 hours
Loo et al.	2007	57	16 to 18	94.6 (11.1)	131	16 to 18	94.6 (11.1)	-0.200	Community	Multirater	Mixed	≥24 hours
Newcorn et al.	2001	31	7.76 (0.77)	101.35 (13.86)	117	7.76 (0.77)	101.35 (13.86)	-0.141	Clinical	Multirater	ADHD-C	<24hours
Poissant et al.	2016	7	11.00 (1.51)	103.00 (8.00)	16	10.00 (2.30)	107.00 (15.30)	1.284**	Clinical	Multirater	Mixed	≥24 hours
Rucklidge	2006	25	14.94 (1.22)	97.12 (11.16)	24	14.78 (0.99)	98.46 (14.13)	-0.233	Clinical	Multirater	Mixed	≥24 hours
Seymour et al.	2016	40	10.40 (1.80)	110.10 (13.80)	41	10.70 (2.10)	107.50 (12.60)	0.000	Community	Single Rater	Mixed	≥24 hours
Siedman et al.	2005	101	12.50 (2.60)	105.70 (11.50)	103	12.70 (2.60)	109.00 (13.50)	0.000	Clinical	Single Rater	Mixed	<24hours

Table 1 (continued)

Effect Sizes and Moderators for Attentional Control

Study	Year	Girls *N*	Age *M*(*SD*)	FSIQ *M*(*SD*)	Boys *N*	Age *M*(*SD*)	FSIQ *M*(*SD*)	Cohen's (*d*)	Referral Source	Diagnostic Raters	Subtype	Medication Washout
Yang et al.	2014	21	8.05 (1.40)	97.10 (7.63)	21	8.05 (1.40)	98.00 (11.64)	-0.011	Clinical	Clinical	ADHD-C	≥24 hours
Weighted mean effect size								-0.071				
95% confidence interval								-0.242 to 0.100				
Q Homogeneity Index								9.929				

Note: Positive effect sizes indicate greater impairment for boys with ADHD; Negative effect sizes indicate greater impairment for girls with ADHD; N/R = Information not reported; FSIQ = Full-Scale Intelligence Quotient; * $p < .05$; ** $p \leq .01$

Table 2

Effect Sizes and Moderators for Inhibition

Study	Year	Girls *N*	Age *M*(*SD*)	FSIQ *M*(*SD*)	Boys *N*	Age *M*(*SD*)	FSIQ *M*(*SD*)	Cohen's (*d*)	Referral Source	Diagnostic Raters	Subtype	Medication Washout
Breen	1989	13	6 to 11	N/R	13	6 to 11	N/R	-0.138	Clinical	Multirater	Mixed	≥24 hours
	1989	17	8.20	N/R	37	8.10	N/R		Clinical	Multirater	ADHD-H/I	≥24 hours
Horn et al.			(1.30)			(1.50)		0.096				
	2007	57	16 to 18	94.6	131	16 to 18	94.6		Community	Multirater	Mixed	≥24 hours
Loo et al.				(11.1)			(11.1)	0.000				
Newcorn et	2001	31	7.76	101.35	117	7.76	101.35		Clinical	Multirater	ADHD-C	<24hours
al.			(0.77)	(13.86)		(0.77)	(13.86)	0.229				
	2016	7	11.00	103.00	16	10.00	107.00		Clinical	Multirater	Mixed	≥24 hours
Poissant et al.			(1.51)	(8.00)		(2.30)	(15.30)	-.571				
	2006	25	14.94	97.12	24	14.78	98.46		Clinical	Multirater	Mixed	≥24 hours
Rucklidge			(1.22)	(11.16)		(0.99)	(14.13)	-0.209				
	2002	24	15.02[a]	101.67*	35	15.02[a]	101.67*					
Rucklidge &			(1.40)	(10.40)		(1.40)	(10.40)					
Tannock								-0.360	Clinical	Multirater	Mixed	≥24 hours
Seymour et	2016	40	10.40	110.10	41	10.70	107.50		Community	Single	Mixed	≥24 hours
al.			(1.80)	(13.80)		(2.10)	(12.60)	0.869		Rater		
	2005	101	12.50	105.70	103	12.70	109.00		Clinical	Single	Mixed	<24hours
Siedman et al.			(2.60)	(11.50)		(2.60)	(13.50)	0.219		Rater		
Sjowall et al.	2013	56	7 to 13	N/R	46	7 to 13	N/R	0.019	Clinical	Multirater	Mixed	<24hours

Table 2 (continued)

Effect Sizes and Moderators for Inhibition

Study	Year	Girls *N*	Age *M*(*SD*)	FSIQ *M*(*SD*)	Boys *N*	Age *M*(*SD*)	FSIQ *M*(*SD*)	Cohen's (*d*)	Referral Source	Diagnostic Raters	Subtype	Medication Washout
Yang et al.	2014	21	8.05 (1.40)	97.10 (7.63)	21	8.05 (1.40)	98.00 (11.64)	0.377	Clinical	Clinical	ADHD-C	≥24 hours
Weighted mean effect size								-0.102				
95% confidence interval								-0.098 to 0.302				
Q Homogeneity Index								19.880				

Note: Positive effect sizes indicate greater impairment for boys with ADHD; Negative effect sizes indicate greater impairment for girls with ADHD; N/R = Information not reported; FSIQ = Full-Scale Intelligence Quotient; * $p < .05$; ** $p \leq .01$

Table 3

Effect Sizes and Moderators for Set-Shifting

Study	Year	Girls *N*	Age *M(SD)*	FSIQ M(*SD*)	Boys *N*	Age *M*(*SD*)	FSIQ M(*SD*)	Cohen's *d*	Referral Source	Diagnostic Raters	Subtype	Medication Washout
DeShazo	2000	12	11.06 (1.60)	102.58 (12.70)	21	11.06 (1.30)	103.29 (15.57)	-1.347***	Clinical	Multirater	Mixed	≥24 hours
Gunther et al.	2015	89	11.50 (1.60)	97.10 (9.20)	86	11.50 (1.70)	97.20 (11.40)	0.484**	Clinical	Multirater	ADHD-C	≥24 hours
Heydari & Farahani	2016	7	7 to 11	N/R	23	7 to 11	N/R	0.305	Clinical	Multirater	Mixed	≥24 hours
Lockwood et al.	2001	20	9.10 (1.33)	90-129	20	9.31 (1.76)	90-129	0.322	Clinical	Multirater	ADHD-C	≥24 hours
Loo et al.	2007	57	16 to 18	94.6 (11.1)	131	16 to 18	94.6 (11.1)	0.036*	Community	Multirater	Mixed	≥24 hours
O'Brien et al.	2010	26	9.82 (1.18)	108.96 (14.15)	30	10.52(1.32)	107.63 (11.80)	-0.084	Both	Single rater	Mixed	≥24 hours
Roufael et al.	2012	30	9.19 (1.20)	90.22 (5.30)	30	9.18 (1.09)	88.93 (4.72)	0.270	Clinical	Single rater	Mixed	N/R
Rucklidge	2006	25	14.94 (1.22)	97.12 (11.16)	24	14.78 (0.99)	98.46 (14.13)	0.043	Clinical	Multirater	Mixed	≥24 hours
Rucklidge & Tannock	2002	24	15.02* (1.40)	101.67* (10.40)	35	15.02* (1.40)	101.67* (10.40)	0.478	Clinical	Multirater	Mixed	≥24 hours

Table 3 (continued)

Effect Sizes and Moderators for Set-Shifting

Study	Year	Girls *N*	Age *M(SD)*	FSIQ M(*SD*)	Boys *N*	Age *M*(*SD*)	FSIQ M(*SD*)	Cohen's *d*	Referral Source	Diagnostic Raters	Subtype	Medication Washout
Siedman et al.	2005	101	12.50 (2.60)	105.70 (11.50)	103	12.70 (2.60)	109.00 (13.50)	-0.033	Clinical	Single rater	Mixed	<24hours
Sjowall et al.	2013	56	7 to 13	N/R	46	7 to 13	N/R	0.049	Clinical	Multirater	Mixed	<24hours
Skogli et al.	2013	37	12.45 (2.09)	96.40 (15.50)	43	11.60 (1.93)	94.30 (13.20)	0.216	Clinical	Multirater	Mixed	≥24 hours
Wodka et al.	2008	22	11.00 (2.10)	110.80 (14.10)	32	12.40 (2.10)	107.80 (13.50)	0.313	Both	Single rater	Mixed	≥24 hours
Weighted mean effect size								0.157				
95% confidence interval								-0.031 to 0.344				
Q Homogeneity Index								9.929				

Note: Positive effect sizes indicate greater impairment for boys with ADHD; Negative effect sizes indicate greater impairment for girls with ADHD; N/R = Information not reported; FSIQ = Full-Scale Intelligence Quotient; * $p < .05$; ** $p \leq .01$; *** $p \leq .001$

Table 4

Effect Sizes and Moderators for Planning/Organizaiton

Study	Year	Girls *N*	Age *M(SD)*	FSIQ M(*SD*)	Boys *N*	Age *M*(*SD*)	FSIQ M(*SD*)	Cohen's *d*	Referral Source	Diagnostic Raters	Subtype	Medication Washout
DeShazo	2000	12	11.06 (1.60)	102.58 (12.70)	21	11.06 (1.30)	103.29 (15.57)	0.404	Clinical	Multirater	Mixed	≥24 hours
Lockwood et al.	2001	20	9.10 (1.33)	90-129	20	9.31 (1.76)	90-129	-0.684*	Clinical	Multirater	ADHD-C	≥24 hours
O'Brien et al.	2010	26	9.82 (1.18)	108.96 (14.15)	30	10.52(1.32)	107.63 (11.80)	0.090	Both	Single rater	Mixed	≥24 hours
Siedman et al.	2005	101	12.50 (2.60)	105.70 (11.50)	103	12.70 (2.60)	109.00 (13.50)	0.026	Clinical	Single rater	Mixed	<24hours
Skogli et al.	2013	37	12.45 (2.09)	96.40 (15.50)	43	11.60 (1.93)	94.30 (13.20)	0.312	Clinical	Multirater	Mixed	≥24 hours
Wodka et al.	2008	22	11.0 (2.1)	110.80 (14.10)	32	12.4 (2.1)	107.80 (13.5)	-0.211	Both	Single rater	Mixed	≥24 hours
Weighted mean effect size								0.009				
95% confidence interval								-0.246 to 0.264				
Q Homogeneity Index								8.214				

Note: Positive effect sizes indicate greater impairment for boys with ADHD; Negative effect sizes indicate greater impairment for girls with ADHD; N/R = Information not reported; FSIQ = Full-Scale Intelligence Quotient; * $p < .05$; ** $p \leq .01$

Table 5
Effect Sizes and Moderators for Phonological Working Memory

Study	Year	Girls *N*	Age *M(SD)*	FSIQ M(*SD*)	Boys *N*	Age *M*(*SD*)	FSIQ M(*SD*)	Cohen's *d*	Referral Source	Diagnostic Raters	Subtype	Medication Washout
Boseck et al.	2016	42	10.10 (2.90)	N/R	105	10.10 (3.40)	N/R	0.192	Clinical	N/R	N/R	<24 hours
Lockwood et al.	2001	20	9.10 (1.33)	90-129	20	9.31 (1.76)	90-129	0.256	Clinical	Multirater	ADHD-C	≥24 hours
O'Brien et al.	2010	26	9.82 (1.18)	108.96 (14.15)	30	10.52(1.32)	107.63 (11.80)	-0.028	Both	Single rater	Mixed	≥24 hours
Skogli et al.	2013	37	12.45 (2.09)	96.40 (15.50)	43	11.60 (1.93)	94.30 (13.20)	0.145	Clinical	Multirater	Mixed	≥24 hours
Weighted mean effect size								0.146				
95% confidence interval								-0.082 to 0.375				
Q Homogenity Index								0.602				

Note: Positive effect sizes indicate greater impairment for boys with ADHD; Negative effect sizes indicate greater impairment for girls with ADHD; N/R = Information not reported; FSIQ = Full-Scale Intelligence Quotient; * $p < .05$; ** $p \leq .01$

Appendix B

Figures

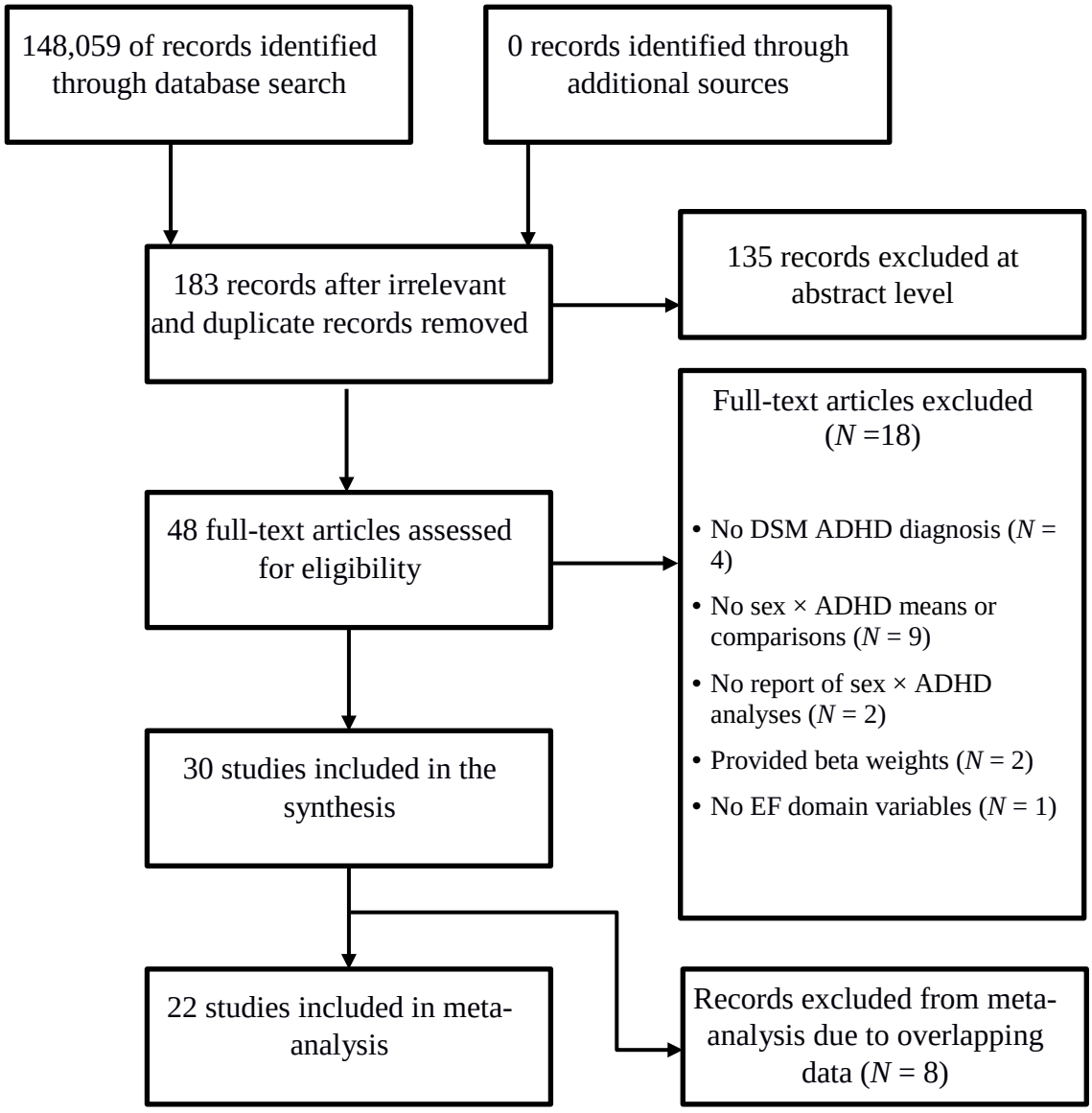


Figure 1. *Flowchart of Included Studies*

Appendix C

Codebook Manual

Codebook Manual

This manual provides instructions and definitions needed for coding and reporting data from each study. A PDF for each study is located in the EndNote Library for ADHD Girls & Boys EF. Extract the appropriate data from the PDF and enter it into your copy of the Excel Coding Scheme worksheet. Data to be coded falls into two categories: study characteristics and participant characteristics. In the Excel Coding Scheme Worksheet, studies are listed by row in alphabetical order by author. Data codes are listed in the top row of each column. Do not type any information into the cells that does not adhere to the coding scheme. If information for a particular code is not available for a study, enter '9999.' If coding information is present for a given study but it does not adhere to definitions of the coding scheme, enter '7777' for subsequent redefining of the coding scheme.

I. Study Characteristics

A. Name

- Enter the last names of each author according to authorship listed in article

B. Date

- Enter the year the journal article was published.

C. Referral Source

- Was the sample pulled from the community (schools, fliers, other advertisements), or were they children recruited from a clinic? Code '1' if the ADHD sample was recruited from the community. Code '2' if the sample was recruited from clinics or psychiatric referrals.

D. Diagnostic raters

- How were diagnoses made? Did parent and/or teachers complete rating scales on the child's behavior? Did a referring health professional previously diagnosis the child with ADHD? Code '2' if *at least two* of the following were used to make a diagnosis: parent ratings, teacher ratings, semi-structured interview conducted by the researchers, and/or previous diagnosis from a psychologist, psychiatrist, or physician. Code '1' if fewer than two of the aforementioned were used to make an ADHD diagnosis.

E. Subtype

- What subtypes of ADHD were included in the sample? Code '1' for ADHD-Combined. Code '2' for ADHD-Predominately Inattentive. Code '3' if the sample included multiple subtypes, if the sample did not (or was conducted prior to subtype stratification), or if the ADHD-I or ADHD-H/I were combined with the ADHD-C group.

F. Medication

- Were children taking stimulant medication at the time of testing? Code '1' if children were tasked to discontinue medication at least 24 hours prior to testing or if the sample was medication naïve. Code '2' if children were not asked to discontinue medication at least 24 hours in advance or if children were retained in the sample who forgot/whose parents were unwilling to discontinue medication.

II. Sample Characteristics

A. ADHD girls sample size
- List the number of girls with ADHD included in the analyses. Do not include numbers of girls with ADHD and a comorbid condition if this is a separate group of analyses. If ADHD groups are separated by subtypes, provide data for the ADHD, Combined subtype.

B. ADHD girls mean age
- List the mean age of ADHD girls included in the analyses. If no mean is available, provide the range of ages included. If mean age is provided for total ADHD sample or total sample of children, enter that mean.

C. *SD* for ADHD girls' ages
- Enter the *SD* of ADHD girls included in the analyses. If only age range is available, enter the *SD* of the range. If *SD* age is provided for total ADHD sample or total sample of children, enter that *SD*.

D. ADHD Girls FSIQ
- Enter the Full-Scale IQ for the ADHD girls included in the analyses.

E. ADHD Girls FSIQ *SD*
- Enter the *SD* or confidence interval for the FSIQ for ADHD Girls.

F. ADHD Boys sample size
- List the number of boys with ADHD included in the analyses. Do not include numbers of boys with ADHD and a comorbid condition if this is a separate group of analyses. If ADHD groups are separated by subtypes, provide data for the ADHD, Combined subtype.

G. ADHD Boys' mean age
- List the mean age of ADHD boys included in the analyses. If no mean is available, provide the range of ages included. If mean age is provided for total ADHD sample or total sample of children, enter that mean.

H. *SD* for ADHD boys' ages
- Enter the *SD* of ADHD boys included in the analyses. If only age range is available, enter the *SD* of the range. If *SD* age is provided for total ADHD sample or total sample of children, enter that *SD*.

I. ADHD Boys FSIQ
- Enter the Full-Scale IQ for the ADHD boys included in the analyses.

J. ADHD Boys FSIQ *SD*

- Enter the *SD* or confidence interval for the FSIQ for ADHD Boys.

Vita

Jenna Elizabeth Gilmore was born and raised in Edmond, Oklahoma. She received her Bachelor of Arts degree in Psychology from Oklahoma State University in 2012 and began her doctoral training in Clinical Psychology at the University of Tennessee – Knoxville in 2014. Her research interests include the neuropsychological correlates of developmental psychopathology and the broad intersection of personality and neuropsychological functioning.