EXAMINATION OF A NOVEL UNSTRUCTURED PERFORMANCE-BASED TASK OF EXECUTIVE FUNCTION IN CHILDREN WITH ATTENTION DEFICIT HYPERACTIVITY DISORDER AND A COMMUNITY SAMPLE OF TYPICALLY DEVELOPING CHILDREN

JUSTINE MAYA LEDOCHOWSKI

A THESIS SUBMITTED TO THE FACULTY OF GRADUATE STUDIES IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF ARTS

GRADUATE PROGRAM IN PSYCHOLOGY

YORK UNIVERSITY

TORONTO, ONTARIO

AUGUST 2016

Abstract

Executive functions enable problem solving and goal attainment. EF have been assessed with performance-based measures and rating scales. Research has shown a lack of association between these two methods. One framework used to understand this difference is the structure provided on performance-based measures and not provided on rating scales. This study investigated the role of structure through examining a novel Unstructured Performance-based Task (UPT). Eighty children between 8-12 (38 with ADHD, 42 typically developing, M_{age}= 9.56, SD= 1.29) and their parents participated. Significant associations emerged between the UPT, performance-based measures and rating scale. Performance-based measures and the rating scale significantly predicted UPT performance. Group differences were found across all measures. The UPT significantly predicted ADHD status when entered with performance-based measures, but not the rating scale. Results suggest UPT may be a promising measure to assess EF related difficulties in ADHD and provide a picture of children's behaviours in unstructured environments.

Acknowledgements

First and foremost, I would like to thank my supervisor, Dr. Maggie Toplak, for her support, guidance, and knowledge throughout this project. I'd like to thank my second committee member Dr. Brendan Andrade for his very helpful feedback as well as his aid with recruitment. I would also like to thank the members of the Toplak Research Lab for their help and support throughout this project and preparation for my defense. A special thank you goes to Alexandra Basile for allowing me to join her dissertation project to collect my data and Rachael Lyon who worked diligently testing participants for this project. I also extend thanks to my oral exam members Drs. Caroline Davis and Mary Desrocher for their thought-provoking and insightful feedback. I would like to mention the children and families who participated in this study, without them this research project would not be possible and I thank them for their time. Finally, I would like to thank my family and friends for their love and support throughout my education.

Table of Contents

List of Tables

List of Appendices

List of Acronyms

ADHD= Attention Deficit Hyperactivity Disorder

BDEFS-CA= Barkley Deficits in Executive Function Scale – Children and Adolescents Short Form

EF= Executive Functions

IQ= Intelligence quotient

TD= Typically developing

TMT= Trail-Making test

UPT= Unstructured Performance-based Task

Examination of a Novel Unstructured Performance-Based Task of Executive Function in Children with Attention Deficit Hyperactivity Disorder and a Community Sample of Typically Developing Children

Executive Functions (EFs) are a set of cognitive skills that enable problem-solving and goal-directed behaviour (Lezak, 1995; Welsh & Pennington, 1998). EFs have traditionally been assessed using two methods: performance-based measures and rating scales. Although both of these forms of measurement were designed to assess EFs, past research has shown a lack of convergence of information between performance-based measures and rating scales (Bodnar, Prahme, Cutting, Denckla, & Mahone, 2007; Gray, Fettes, Woltering, Mawjee, and Tannock, 2015; Mahone et al., 2002; McAuley, Chen, Goos, Schachar, & Crosbie, 2010; Toplak, West, & Stanovich, 2013). The lack of convergence between these forms of measurement may be attributed to the design and purpose of these two forms of measurement. Rating scales were designed to be ecologically valid, that is to provide information about an individual's functioning in everyday, real-word circumstances, such as at home, school, or in the community (Gregory, 2011; Holtz, 2011). In contrast, performance-based measures are completed during structured one-to-one testing in a laboratory. Thus, these forms of measurement may be assessing different but related aspects of EF (Toplak et al., 2013). Rating scales may be tapping into information about how the individual's EF manifests in everyday activities, whereas performance-based measures may be a better representative of functioning during structured activities, in which less self-regulation on the part of the examinee is required. In order to examine the role of structure in performance-based tasks as an explanation for the discordance between EF rating scales and performance-based tasks a novel Unstructured Performance-Based Task (UPT), in which the structure provided by the examiner is minimized and left somewhat open to interpretation by the

examinee was utilized. Furthermore, past research has shown reliable differences in EF between individuals with Attention Deficit Hyperactivity Disorder (ADHD) and non-clinical controls, using both performance-based tasks and rating scales (Barkley & Fischer, 2011; Barkley & Murphy, 2010; Biederman et al., 2007; Mahone et al., 2002; Mahone & Hoffman, 2007; Nigg, Willcutt, Doyle, & Sonuga-Barke, 2005; Toplak, Bucciarelli, Jain, & Tannock, 2005; Scheres et al., 2004; Willcutt, Doyle, Nigg, Faraone, & Pennington, 2005). As such, this study examined the UPT in children with ADHD and typically developing children (TD). There were two major goals in this study. First, to examine the role of structure in performance-based measures by investigating associations between performance-based tasks, a rating scale, and the novel Unstructured Performance-Based Task (UPT). Second, to determine the whether the UPT may be diagnostic in identifying EF differences in children with ADHD and typically developing children (TD).

Plan of Introduction

The introduction will begin with a brief review of EF. Next, I will summarize measurement of EF, in particular performance-based tasks and rating scales. I will then review the literature concerning associations between performance-based tasks and rating scales. Next, I will provide possible explanations for the lack of convergence between EF as measured by performance-based tasks and rating scales. I will then discuss findings concerning the role of EF in ADHD. Finally, I will introduce the current study, providing a brief introduction of the UPT, as well as research objectives, and hypotheses.

Executive Function

Executive Functions (EFs) are a set of cognitive skills that enable problem-solving and goal-directed behaviour (Lezak, 1995; Welsh & Pennington, 1998). Examples of executive

functions include inhibition, planning, self-monitoring, self-regulation, working memory, response inhibition, and set-shifting (Lezak, 1995; Pennington & Ozonoff, 1996). EFs emerge early in life, continue to develop until adolescence or early adulthood, and begin to decline in adulthood, with relative stability within individuals over the course of development (Miyake & Friedman, 2012; Romine & Reynolds, 2005; Zelazo, Craik, & Booth, 2004). In childhood EF development is associated with increases in cognitive control and efficiency of cognitive processes (Carlson, Zelazo, & Faja, 2012). Additionally, modest to strong associations between EF and intellectual ability have been noted (Arffa, 2007; Brydges, Reid, Fox, & Anderson, 2012).

Measurement of Executive Function

EF has typically been assessed in two ways. First, EF has been assessed through structured performance-based tasks that are administered in a standardized environment by an examiner. These tasks are administered according to standardized procedures, meaning that stimuli are presented in the same manner to every examinee and tasks are completed in the same fashion by every participant, often in a limited time frame (i.e., lasting between 5-30 minutes) (Toplak et al., 2013). Performance-based tasks are typically standardized and have norms that the examiner can use to compare the examinee's performance to a sample of other similarly aged peers. As such, the examiner establishes standard testing conditions in order to ensure each examinee is exposed to nearly identical testing conditions so that meaningful information can be drawn from these norms. To this end, performance-based measures are structured, examiners present highly detailed instructions, specific prompts and feedback. These tasks usually assess accuracy, response time, or speeded responding (Toplak, et al., 2013). Examples of such tests include the Stroop Colour-Word Test (Golden, 1978), Trail-Making test (Reitan, 1971),

Wisconsin Card Sorting test (Heaton, Chelune, Talley, Kay, & Curtis, 1993) and tests of verbal fluency (Strauss, Sherman, & Spring, 2006).

The second commonly employed method to assess EF is behavioural rating scales. These scales were developed to be an ecologically valid measure of children's behaviour in everyday situations (Roth, Isquith, & Gioia, 2005). Ratings scales present a list of behaviours and ask the reporter (e.g., parent, teacher, or self) to indicate how often the child performs these behaviours in everyday situations. Examples rating scales include the Behaviour Inventory of Executive Function (BRIEF; Gioia, Isquith, Guy, & Kenworthy, 2000), the Childhood Executive Functioning Inventory (CHEXI; Thorrell, Eninger, Brocki, & Bohlin, 2010) and the Barkley Deficits in Executive Functioning for Children and Adolescents (BDEFS-CA; Barkley, 2012). The BRIEF consists of questions that assess EF domains of behavioural inhibition, shifting, emotional control, initiation, working memory, planning/organization, and monitoring. Similarly, the BDEFS-CA asks questions about daily activities pertaining to the EF domains of time management, organization, problem- solving, self-restraint, self-motivation, and self-regulation of emotions. The CHEXI consists of questions about children's working memory and inhibition.

Associations Between Performance-based Measures and Rating Scales to Assess Executive Function

According to the principles of convergent validity it is expected that measures that assess the same construct will be highly correlated with one another (Gregory, 2011). Thus, as both performance-based measures and rating scales are intended to assess EF, it would be expected that there would be strong and positive associations between these measurement approaches. However, past research has shown a surprising lack of association between performance-based measures and rating scales assessing EF. Mahone et al. (2002) investigated the convergence and

divergence of the BRIEF and performance-based measures of inhibition, planning, and oral fluency in children with ADHD and/or Tourette's disorder as well as typically developing children. Results showed low to moderate correlations between performance-based measures and the rating scale, however none of the correlations were significant. Bodnar et al. (2007) compared the BRIEF and performance-based measures of inhibitory control in mixed outpatient sample of children with various psychiatric and neurological conditions. In both groups few correlations between the rating scale and performance-based measures were statistically significant, and of those that were, the correlations were small. Similarly, McAuley et al. (2010) investigated the associations between the BRIEF and performance-based measures assessing inhibition and working memory in youth referred for attention, learning, and/or behavioural problems as well as healthy controls and found no significant associations. Additionally, Gray et al. (2015) investigated symptoms and impairment in a sample of post-secondary students with ADHD. Students reported clinically significant symptoms of ADHD and EF difficulties in everyday life in a semi-structured phone interview. In contrast, the same students performed in the average range on performance-based measures of EF. Toplak et al. (2013) provided further evidence for the lack of convergence between performance-based measures and rating through a review of 20 studies examining associations between measures among various populations (e.g., child, adult, non-clinical, and different clinical populations including psychiatric and neurological conditions). Results showed that only 24% of relevant correlations reported in the studies was statistically significant, and that the median magnitude of the significant correlations was small.

Toplak et al., (2013) suggest that rating scales and performance-based measures index different but related mental constructs. Specifically, the authors suggest that performance-based

measures assessing EF provide information about the efficiency of cognitive processes that can be recruited during behavioural control (e.g., inhibition), whereas rating scales of EF provide information about how EF manifests in achieving goals in everyday activities. As such, performance-based measures and rating scales should not be interpreted as equivalent or interchangeable measures of the same construct, but instead regarded as providing independent information about an individual's EF.

Proposed Reasons for Discordance Between Performance-based Measures and Rating Scales: Ecological Validity and Structure

There are several different reasons why rating scales and performance-based measures seem to display low associations. As previously mentioned performance-based measures assessing EF are highly structured and administered in a standardized environment with explicit instructions provided by the examiner. This approach greatly alleviates the self-regulation demands of the examinee (Clark, Prior, & Kinsella, 2000). The role of children's EF abilities on task performance may be less necessary, given that the structure and organization of the task are provided by the examiner rather than demanded of the examinee (Salthouse, Atkinson, & Berish, 2003). As such, it is not necessary for the examinee to make decisions regarding the task goal, strategies to employ, or whether ongoing behaviour needs modification (Clark et al., 2003). Thus, the testing environment in which performance-based measures are completed do not reflect EF demands in everyday life. For example, whilst doing independent work at school a child must be able to plan how to approach the work, self-regulate their behaviour to do the work and not be distracted by preferred activities, as well as make decisions about what to do if their chosen approach does not work (e.g., if they do not know how to solve a problem on the work sheet they can skip it, mark it for homework later, or ask the teacher for help). It is important to

note that there exist a number of performance-based measures in the literature designed to minimize the structure imposed by the examiner. Though efforts have been made to minimize the involvement of the examiner in guiding the examinee's actions and providing a less strictly structured setting for task completion these measures still impose several structured constructs. See Appendix A for a review of these measures.

In contrast, rating scales were designed in order to capture executive dysfunction that may arise in daily activities, and as such, may provide an ecologically valid measure of EF (Roth et al., 2005). Ecologically valid measures are designed to provide information about a person's functioning in real-world circumstances, such as at home, work, school, or in the community (Gregory, 2011; Holtz, 2011). For this purpose, rating scales consist of questions about the child's behaviour in everyday, real-life situations over roughly the past 6 months. Thus, the information collected via rating scales does not impose structure on the examinee as do performance-based measures. An important assumption underlying rating scales is that behaviours assessed by these scales are in fact related to the EF processes that are assessed by performance-based measures (Toplak et al., 2013).

An analogy can be made to typical performance situations and optimal performance situations. Typical performance situations are those in which there are no instructions given on how to maximize performance and it is up to the participant to interpret the task, where as in optimal performance situations explicit goals are provided as to how to maximize performance and task interpretation is constrained (Toplak et al., 2013). Rating scales would reflect behaviour that is more closely related to typical performance, whereas due to their structure, performance-based measures provide information about an individual's functioning in optimal performance conditions.

The role of structure in performance-based measures has been investigated in children with ADHD. Behavioural observation of children with hyperactive behaviour indicates calm, attentive, and minimal observable EF deficit during performance-based measures (Mahone & Hoffman, 2007). Furthermore, in the context of a highly structured situation with explicit rules and short-term consequences for inappropriate behaviour children with hyperactive behaviour can appear calm and attentive while completing performance-based measures assessing visual and auditory attention (Draeger, Prior, & Sanson, 1986). In contrast, children with ADHD demonstrate difficulty in everyday situations, such as waiting turns, free play, and in loosely structured classroom activities (Clark et al., 2003; Imeraj et al., 2016). Thus, the effect of structure provided in performance-based measures compared to that provided in rating scales is especially notable in this population.

Executive Function and Attention Deficit Hyperactivity Disorder

ADHD is a neurodevelopmental disorder that affects an estimated 5% of children (American Psychiatric Association [APA], 2013; Polanczyk & Rohde, 2007). It is characterized by hyperactivity, inattention, and impulsive behaviour (APA, 2013). Inattentive symptoms include behaviours such as daydreaming, distractibility, and difficulty focusing on a single task for a prolonged period of time (Biederman, 2005). Hyperactive and impulsive symptoms include behaviours such as fidgeting, difficulty waiting one's turn, and discomfort being still for extended periods of time or feeling restless (APA, 2013). EF deficits have been proposed to be important neuropsychological correlates of ADHD (Pennington & Ozonoff, 1996). Reliable differences in EF have been found between individuals with ADHD and non-clinical controls, using performance-based measures and rating scales. On performance-based measures of EF including set-shifting, working memory, inhibition, and planning both children and adults with

ADHD perform worse compared to non-clinical controls (Nigg et al., 2005; Scheres et al., 2004). Further, a meta-analysis by Willcutt, Doyle, Nigg, Faraone, and Pennington (2005) showed that youth with ADHD have significant performance deficits on all EF performance-based tasks with the strongest and most consistent deficits in response inhibition, working memory, vigilance, and planning. Similarly, when using rating scales to examine EF deficits, individuals with ADHD show more difficulties in everyday behaviours compared to non-clinical controls (Barkley & Fischer, 2011; Barkley & Murphy, 2010; Biederman et al., 2007; Mahone et al., 2002; Mahone & Hoffman, 2007; Toplak, Bucciarelli, Jain, & Tannock, 2005). Importantly, EF deficits are associated with difficulties in behavioural, social, and academic functioning of children with ADHD at home and at school (Clark et al., 2002; Willcutt et al., 2005). Furthermore, EF deficits in individuals with ADHD have lasting effects into adulthood, negatively affecting socioeconomic status, educational attainment, and occupational attainment in adolescents and adults (Barkley & Murphy, 2010; Biederman et al., 2004).

The Current Study

The purpose of the current study is to examine the role of structure in performance-based measuress as an explanation for the discordance between EF rating scales and performance-based tasks. For this purpose, a novel Unstructured Performance-Based Task (UPT; See Appendix B), which minimizes the structure imposed by the examiner, was utilized.

Development and rationale of the Unstructured Performance-based Task. The premise underlying development of the UPT is that by limiting the structure in a performance-based task it may provide the examiner with greater understanding of how the examinee functions in daily activities where structure is not imposed and they must use self-regulation. It consists of 42 questions presented in scattered order on a sheet of 11x17 inch piece of paper. The

random, scattered order of the items of was chosen so that examinees cannot complete each item systematically and rather must find a way to complete it themselves. Thus, structure is not provided for the examinee in this task. Questions were selected from the domains of math, reading, and general knowledge, as well as rote copying tasks. Items were designed to be simple for children aged 8-12 to complete. Items are not meant to rely heavily on knowledge or processing. However, to account for any individual differences in what is considered a "hard item" the option to circle items that an examinee did not know the answer to was given. Thus, examiners knew the item was attended to by the examinee, in contrast to items left blank, which were not attended to by the child. Furthermore, instructions accompanying the task are brief and do not contain information about the best way to complete the task, again leaving it up to the examinee to determine the best way to approach the task (see Appendix C for a copy of instructions). Additionally, because past research has shown that children with ADHD can present as less hyperactive and more attentive during one-to-one performance-based testing, it is of interest to explore whether the structure of performance-based tasks is affecting children's hyperactive behaviour. For this purpose, an accompanying Behavioural Checklist that the examiner completes while the child is completing the UPT was designed. The UPT and accompanying Behavioural Checklist are further described in the Methods section.

Research Objectives

This study had three objectives. First, to examine the pattern of associations between rating scales, performance-based measures, and the UPT as well as the associations between the UPT, age, and intellectual ability. Second, to compare EF as measured by rating scales, performance-based measures, and the UPT in children with ADHD and TD children. Third, to determine whether the UPT significantly predicted ADHD status. ADHD diagnosis will be

confirmed during testing with a structured clinical interview administered to the caregiver. Children will complete performance-based measures of EF and parents will complete a rating scale assessing EF. In order to address these research objectives there were five hypotheses.

Hypotheses

Hypothesis 1. The UPT will be significantly correlated with both the performance-based and rating measures of EF. Using regression analyses, it is expected that performance-based measures and rating scales will both be significant predictors of performance on the UPT

Hypothesis 2. Performance on the UPT will be significantly correlated with age and intellectual ability as both have been shown to be associated in the literature (Arffa, 2007; Brydges et al., 2012; Romine & Reynolds, 2005; Miyake & Friedman, 2012; Zelazo et al., 2004).

Hypothesis 3. There will be group differences between children with ADHD and TD children such that children with ADHD will show greater difficulties on the EF ratings compared to TD children, as has been shown in the literature (Barkley & Fischer, 2011; Barkley & Murphy, 2010; Biederman et al., 2007; Mahone et al., 2002; Mahone & Hoffman, 2007; Toplak et al., 2005), children with ADHD will show poorer performance on performance-based measures of EF compared to TD children, as has been shown in the literature (Nigg et al., 2005; Scheres et al., 2004; Willcutt et al., 2005), and children with ADHD will show poorer performance on the UPT compared to TD children, as this measure will tap into EF, which has been shown to be impaired in children with ADHD.

Hypothesis 4. There will be group differences such that children with ADHD will exhibit more fidgety behaviours, off-task behaviours, and verbalizations as indicated in the accompanying UPT Behavioural Checklist compared to TD children during UPT task completion.

Hypothesis 5. The UPT will significantly predict ADHD status, even after entering performance-based measures and rating scales of EF.

Methods

Participants

Eighty-six children and their caregivers from the Greater Toronto Area participated in the study. However, six children had to be excluded for various reasons (discussed below). Thus, a total of 80 children between the ages 8-12 (*M*= 9.56 years, *SD*= 1.29, 26 females) and their caregivers were included in the analyses. The sample consisted of two groups of children, a clinical sample of children with a diagnosis of ADHD and a typically developing (TD) control sample of children. There were no significant differences in age (p=.817) or gender (p= .521) between groups. Inclusion criteria were as follows: (1) children between the ages of 8-12 years, (2) a prior diagnosis of ADHD and meeting diagnostic criteria for ADHD on the Computerized Diagnostic Interview Schedule for Children – Parent Version (C-DISC; Fisher, Lucas, Lucas, Sarsfield, & Shaffer, 2006) to be in the clinical group, (3) absence a prior diagnosis of ADHD and must not meet diagnostic criteria for ADHD on the C-DISC to be in the TD group, and, (4) borderline or above intellectual functioning (i.e., IQ $\geq$ 70) and without an Autism Spectrum Disorder (ASD). Children meeting these criteria were invited to participate in the study.

Exclusions. Six participants were excluded from the study. Five participants were excluded from the clinical group. Four participants were excluded because they did not meet inclusion criteria as they did not meet ADHD diagnostic criteria on the CDISC. One participant was excluded because his IQ score was below the inclusion criteria of 70. One participant was excluded from the TD group because he met criteria for ADHD on the CDISC.

Clinical group. The clinical ADHD group was recruited from an outpatient mental health service, private psychological practices, from a previous study conducted at York University, and online at the Canadian ADHD Resource Alliance (CADDRA) and the Centre for ADHD Awareness, Canada (CADDAC). Recruitment was completed over email and telephone. Prior ADHD diagnosis was established and later confirmed during testing using a structured clinical interview. The clinical group consisted of 11 females and 27 males aged 8-12 (*M*= 9.55 years, *SD*= 1.37). Parents identified half of the children as Caucasian (50%), 31.6% as mixed background, 15.8% as other, and 2.6% as Latin American. Ethnic backgrounds represented among children identified as mixed background included Caribbean, Indochinese, Czech and Slovak, and Aboriginal.

Parents reported half of the children were taking psychotropic medication (50%). All children taking psychotropic medication were taking psychostimulants. One child was taking a selective-serotonin reuptake inhibitor to treat anxiety in addition to a psychostimulant. Additionally, one child was taking homeopathic remedies indicated for concentration. As several distributions were not normal an Independent Samples Mann-Whitney U Test was conducted to determine if there were significant differences between IQ and EF (e.g., Stroop test, Trail Making test) between children who were taking psychotropic medication and those who were not. There were no significant differences in IQ, Trail Making Test Trails B-A, and the Stroop test (Interference Condition – Colour Naming completion time) across the two groups (p = .644; p = .181; and p = .258, respectively).

Ten children (26.3%) were indicated as having a Learning Disorder. Parents who specified what type of learning disorder indicated that one child was diagnosed with dyslexia, one child was diagnosed with dyslexia, dysgraphia, and non-verbal learning disorder, one child

had a specific learning disorder in math, writing, and borderline reading, one child had specific learning disorders in executive functioning and phonological processing, one child was indicated as having impairments in written expression, one child's impairments were indicated as being "related to reading and writing", and one child had impairments with language coding. Two (5.3%) children were indicated as having a Language Impairment. One child was indicated as having consistent speech and language delays since the age of three, and one child was indicated as having speech and language delays prior to the age of six but has since fully developed his speech and language abilities and no longer has delays.

Eighteen children (47.37%) had comorbid Oppositional Defiant Disorder (ODD), one child (2.63%) had comorbid Conduct Disorder (CD), and three children (7.89%) had both comorbid ODD and CD.

Typically developing group. The TD group was recruited from a list of families in the community who have indicated interest in being contacted to participate in research. Recruitment was completed over email and telephone. Absence of prior ADHD diagnosis was established and later confirmed during testing using a structured clinical interview The TD group consisted of 15 females and 27 males aged 8-12 years (M= 9.57, SD= 1.23). Parents identified 78.6% children as Caucasian, 16.7% as having a mixed background, 2.4% as Latin America, and 2.4% as other. Categories represented in children identified as mixed background included Caribbean, African, European, Arab, and East Asian. No parents reported any of the TD children to be taking any psychotropic medication. Two children were on homeopathic remedies that parents indicated as for "defensive purposes". One child (2.4%) was identified as having a Learning Disorder, with impairments in reading and spelling. One (2.4%) child was indicated as having a Language

Impairment, specifically this child had a tongue thrust and pronounced “V’s” as “B’s”. No children in the TD group had ODD or CD.

Parent demographics. In the control group, 39 mothers and three fathers completed the study with their child. The marital status of parents in this group included 38 married parents, three common-law marriages, and one single parent. Level of mother’s education in this group included one mother who completed Grades 9-12, three mothers who completed some post-secondary education, and 37 mothers who completed post-secondary education.

In the clinical group, 32 mothers, three fathers, one grandmother, and two adoptive fathers completed the study with their child. The marital status of parents in this group included 19 married parents, two common-law marriages, 10 single parents, 4 divorced parents, one engaged parent, one separated parent, and one parent who did not indicate their marital status. Level of mother’s education in this group included one mother who completed up to Grade 8, four mothers who completed Grades 9-12, 10 mothers who completed some post-secondary education, and 23 mothers who completed post-secondary education, and mother who did not disclose her highest level of education.

Procedure

This study was embedded within a larger study examining performance calibration. Two examiners met with each child and their parent(s). Examiners were a doctoral level student, master’s level student, and an advanced bachelor’s student. The study was explained to each child and parent separately, and informed consent and assent were obtained. One examiner administered measures to the child, and the other examiner administered measures to the parent and completed the computerized diagnostic interview with the parent(s). Time to complete testing ranged from 90 to 120 minutes; each participant received a small honorarium of $20.

The child measures were counterbalanced so that half of the battery was presented first or last determined by whether the participant had an odd or even identification number. Participants with odd numbers received version A of the battery and participants with even numbers received Version B of the battery. Mann Whitney-U tests were used to compare performance on performance-based measures of EF and the UPT to determine whether there were any differences based on task order. There were no significant differences on the Stroop test (p= .220), the Trail-Making test (p= .919), or the UPT (p= .657).

Measures

Parent measures.

Demographics Questionnaire. Participants completed a demographics questionnaire about themselves and their child. Questions included information about the child such as age, gender, country of birth, ethnicity, and medications. Parents provided information about themselves such as ethnicity, marital status, and socioeconomic indicators such as highest level of education and occupational status.

Computerized Diagnostic Interview Schedule for Children – Parent Version (C-DISC; Fisher et al., 2006). The C-DISC is a computerized structured interview designed to assess DSM-IV psychiatric disorders, symptoms, and level of impairment in children and adolescents aged 6 to 17. Note that the symptoms and cut off points are parallel between versions of the DSM-IV and DSM 5 (i.e., 6 out of 9 symptoms of either inattention or hyperactivity must be met). Trained graduate students administered the Attention/Deficit-Hyperactivity Disorder, the Oppositional Defiant Disorder, and the Conduct Disorder subscales. Parents were first asked whether their child had experienced a specific symptom and then answered follow-up questions

for symptoms that were endorsed. This measure was used to characterize each sample, determine level of impairment on each subscale, and as inclusion and exclusion criteria.

Child Behavior Checklist (CBCL; Achenbach & Rescorla, 2001). The CBCL is a paper-and-pencil checklist completed by parents used to assess various emotional and behavioural problems in children and adolescents aged 6 to 18. Questions are scored on a three point scale ranging from 0= Absent, 1= Occurs Sometimes, and 2= Occurs Often. In order for an item to be endorsed it must have occurred within the last 6 months. This measure has high reliability and validity (Achenback & Rescorla, 2001). The CBCL contains eight syndrome scales: anxious/depressed, depressed, somatic complaints, social problems, thought problems, attention problems, rule-breaking behaviour, and aggressive behaviour. These scales group into the higher factors of internalizing and externalizing. Additionally, there are six DSM-5 oriented scales consistent with the DSM diagnostic categories of: affective problems, anxiety problems, somatic problems, ADHD, oppositional defiant problems, and conduct problems. This measure was used to characterize the sample and to support inclusion/exclusion criteria.

Barkley Deficits in EF Scale – Children and Adolescents Short Form (BDEFS-CA; Barkley, 2012). The BDEFS-CA Short Form is a measure used to assess EF in children and adolescents as it pertains to daily activities. The BDEFS-CA Short Form contains questions pertaining to the domains of time management, organization, problem solving, self-restraint, self-motivation, and self-regulation of emotions. It contains 20 questions, which are rated as 1= Never, 2= Sometimes, 3= Often, and 4= Always. The reporter is asked to answer the questions as they pertain to the last six months. The possible range of scores is 20-80. This measure has been found to be both reliable and valid (Barkley, 2012). Parents completed this measure to characterize the EF of each participant. The EF Summary Score will be used, which is the total

score comprised by summing the answers from all 20 questions, as the dependent variable on this task. Scores were reflected such that higher scores indicate better executive function

Child measures.

Intellectual and executive functioning measures.

Kaufman Brief Intelligence Test, Second Edition (KBIT-2; Kaufman & Kaufman 2004). The KBIT-2 measured crystalized and fluid intelligence. Crystalized intelligence was assessed with two orally presented verbal tasks (receptive and expressive vocabulary tasks) that do not involve reading or spelling, and ask questions pertaining to verbal knowledge and riddles. Fluid intelligence was assessed non-verbally with a matrix-reasoning task. This measure has a high degree of reliability and validity (Kaufman & Kaufman, 2004). The possible range of scores is 40-160. The raw scored of each of these subtests were standardized and summed to create a non-age corrected *z*- score, which was the dependent measure in this study. Higher scores are indicative of higher intellectual ability.

Trail-Making Test (TMT; Reitan, 1971). The TMT provides a measure of set-shifting. Set-shifting is a cognitive task that requires one to display flexibility when there are changing rules/schedules of reinforcement in the environment (Strauss et al., 2006). The task involves two components, both of which have a practice version before beginning. In Part A the participants connected 25 numbered circles in numerical order with a pencil. In Part B the participants were instructed to connect alternating letters and numbers in alpha-numerical order (i.e., 1 to A, A to B, B to 2 and so on). Part B contains 13 numbered and 12 lettered circles. The dependent measure for this task was the completion time for Part B minus the completion time on Part A. Part B required the participants to "set-shift" between letters and numbers. Subtracting Part A from Part B controls for processing speed, thus the remaining time is a measure of sets-shifting

alone. Scores were reflected such that higher scores indicate better set-shifting abilities. One participant had extreme outlying scores, which were imputed with the next most extreme value within their group.

Stroop Colour-Word Test (Golden, 1978). The Stroop test measures interference control, a type of inhibition (Friedman & Miyake, 2004). Interference control is the ability to filter out irrelevant information and select relevant information. There were three conditions, each presented with 48 items arranged in a 6x8 matrix. In the word reading condition participants were presented with 48 words printed in black ink that named four colours (red, blue, green, yellow) and they were asked to read the words as quickly as they could without making any errors. In the colour naming condition participants were presented with 48 patches of colour (red, blue, green, yellow) and asked to name the colours as quickly as they could without making any errors. In the interference condition participants were presented with a 48 words printed in various ink colours (red, blue, green, yellow). In this condition the words were printed in an incongruent ink colour (i.e., the word red was printed in blue ink). Participants were asked to name the colour of the word as quickly as possible without making errors. The last condition is the most difficult because participants must inhibit the dominant response of reading the words. The dependent variable of the Stroop task was the total naming time for the interference condition, minus the total naming time for the colour naming condition, which provides the inhibition score (Strauss et al., 2006). Scores were reflected such that higher scores indicate greater inhibition. Note: for the multiple and logistic regressions a composite performance-based task score was computed from the Stroop and TMT.

Experimental measure.

Unstructured Performance Task (UPT). This novel task was developed with the purpose of assessing performance while minimizing structure imposed by the examiner. Successful performance on this task requires self-direction by the participant, as the examiner provides minimal direction. This task is presented on an 11x17 inch sheet of paper. The sheet contains 42 simple questions from the domains of math, reading, and general knowledge as well as rote copying tasks. Items were designed to be easy for children aged 8 to 12 to complete. However, to account for potential differences in what may be considered "simple" children will be informed they have the option to circle an item if they do not understand or know how to complete it. This circling option will account for the difference between children to attend to the question but do not know how to approach it compared to children who do not attend to the question at all. Questions were presented in a random, unstructured order on the page and are not numbered. Thus, it is up to the participant how to approach this task. The same brief instructions were presented to each participant: "I would like you to complete the following worksheet. If you do not know the answer for any of the problems, just circle it and go on to the next problem. I cannot read any of the questions to you. Just do your very best, and when you are done, please bring the worksheet to me." After providing the instructions the examiner remained unobtrusively in the room and did not provide any further assistance with the task. There was an accompanying Behavioural Checklist the examiner completed while the participant was engaged in the task. See Appendix B for a copy of the task and instructions, Appendix C for a copy of the scoring template, and Appendix D for a copy of the Behavioural Checklist. The completion time was recorded for each participant and the task was discontinued after 10 minutes even if the participant had not completed the task; the participant was not informed that completion time

would be recorded or that the task would be discontinued after 10 minutes. Cronbach's alpha was calculated on the 42 items of the UPT as a measure of internal consistency (α= .94), which indicated that this measure was highly reliable. Participants were scored on total number of UPT Items Correct, total number of UPT Items Incorrect, total number of UPT Items Circled, total number of UPT Items Blank and total UPT Completion time. Four dependent variables were calculated: Total Correct (includes: Items Correct), Total Incorrect (includes: Items Incorrect, Items Blank, Items Circled), Total Complete (includes: Items Correct, Items Circled, and Items Incorrect; excludes: Items Blank), and Total Incomplete (includes: Items Blank). Thus, the first pair of variables account for questions the participant answered correctly or incorrectly. The second pair of variables account for completion or failed completion of items (that is, did not leave blank or left blank). These measures of accuracy and completion were the dependent measures on the UPT, as well as UPT Completion time and frequency of behaviours endorsed on the Behaviour Checklist.

Statistical Analyses

All statistical analyses were conducted using SPSS version 20.0. The significance level for research questions was set at the standard $p < .05$. Normality of each of the variables investigated was tested using the Shapiro-Wilk test of normality as well as visual inspection of histograms and Q-Q plots. Some of the distributions were not normal and therefore nonparametric tests such as Spearman Rank Order Correlations and Independent Samples Mann-Whitney U Tests were employed throughout. All assumptions for multiple and logistic regression were met.

Statistical plan. Correlations between scores and indices within the UPT across the entire sample were examined. Additionally, correlations among age, IQ, performance-based measures,

rating scales, and the UPT were examined across the entire sample. A multiple regression was conducted to determine whether performance-based tasks and rating scales predicted performance on the UPT. Group comparisons were conducted on performance on the UPT, performance-based tasks, rating scales, and UPT Behavioural Checklist. Finally, binomial logistic regressions were conducted to determine whether performance-based tasks, the rating scale, or the UPT entered as significant predictors of group status (i.e., ADHD or TD).

Missing data. In the TD group one parent did not complete the CDISC, and two parents failed to complete one to three questions on the BDEFS-CA. In addition, one parent in the TD group disclosed that her ratings on the CBCL were inaccurate and based on her own high expectations for her children, not based on same aged peers. She advised that these data were exaggerated negatively and were therefore inaccurate. In the clinical group one parent failed to complete the reverse side of the CBCL, and two parents failed to complete 1 to 3 questions on the BDEFS-CA. For all missing or invalid data group means (i.e., TD or clinical) were imputed for analyses. For the participant who was missing the CDISC the CBCL was used to determine inclusion/exclusion criteria.

Results

Group Characteristics

Table 1 displays the characteristics of each group on the IQ measure, performance-based EF tasks, rating scales, and UPT. There were no differences between groups on IQ (p= .240). Group differences on the other measures will be discussed in the following sections.

Table 1

Group Characteristics on Study Measures (Mean[SD], Range)

	ADHD (n= 38)	TD (*n*= 42)
Estimated IQ (age corrected)	105.84(14.71), 72-134	109.64(8.42), 90-129
TMT	115.58(68.88), 14-261	82.10(43.70), 12-223
Stroop	52.21(23.98), 20-106	41.29(15.60), 1-74
BDEFS-CA	52.33(10.87), 32-73	29.32(6.65), 21-47
UPT Correct	32(8.84), 9-42	38.05(3.33), 21-41
UPT Incorrect	2.37(1.88), 0-7	2.40(1.45), 0-6
UPT Circled	2.92(5.81), 0-23	.55(1.74), 0-10
UPT Blank	4.71(6.64), 0-29	.95(1.29), 0-7
UPT Total Correct	32(8.84), 0-42	38.05(3.33), 21-41
UPT Total Complete	37.29(6.64), 13-42	41(1.36), 35-42
UPT Completion Time (s)	414.18(140.17), 180-600	335.48(96.51), 131-533

Note: IQ= intelligence quotient, TMT= Trail-Making test, BDEFS-CA= Barkley Deficits in Executive Function Scale – Children and Adolescents Short Form, UPT= Unstructured Performance-based Task

Associations Between Scores of the Unstructured Performance-Based Task

Spearman's rank order correlations among UPT scores can be found in Table 2. Based on these associations as well as the design and development of the UPT the following "UPT Totals" were created, UPT Total Correct (includes: Items Correct), UPT Total Incorrect (includes: Items Incorrect, Items Blank, Items Circled), UPT Total Complete (includes: Items Correct, Items Circled, and Items Incorrect) and UPT Total Blank (includes: Items Blank). These will be used

to investigate associations between the UPT, performance-based tasks, and rating scales.

Table 2

Spearman Correlations Among the UPT Scores Across the Entire Sample

	UPT Correct	UPT Incorrect	UPT Circled	UPT Blank
UPT Correct	-	**-.50****	**-.56****	**-.75****
UPT Incorrect		-	-.02	.20
UPT Circled			-	**.27***
UPT Blank				-

*p =.01; **p <.001
Note: UPT= Unstructured Performance-based Task

Associations Between the UPT and Age, Intellectual Ability, Performance-Based Tasks of Executive Function, and Rating Scale of Executive Function

Spearman's rank order correlations between the UPT with age, intellectual ability, performance-based tasks, and rating scales, can be found in Table 3. Significant small to moderate associations were seen between the UPT and age, intellectual ability, the Stroop task, the TMT, and the BDEFS-CA. In addition, there were no significant associations between the BDEFS-CA and the Stroop test (p= .266), or the TMT (p= .600)

Table 3

Spearman Correlations Between the UPT Totals and Age, IQ, Performance-Based Tasks, and Rating Scale in the Total Sample

	Age (months)	**Intellectual Ability (raw score)**	**Stroop (inhibition)**	**TMT (B-A)**	**BDEFS-CA**
UPT Total Correct	.44***	.60***	.38**	.55***	.36**
UPT Total Incorrect	-.44***	-.60***	-.38**	-.54***	-.37**
UPT Total Complete	.33**	.40***	.23*	.39***	.35**
UPT Total Incomplete	-.33**	-.40***	-.23*	-.39***	-.36**
UPT Completion Time (s)	-.53***	-.45***	-.44***	-.29**	-.24*

*p <.05; **p <.01; ***p < .001
Note: UPT= Unstructured Performance-based Task, TMT= Trail-Making test, BDEFS-CA= Barkley Deficits in Executive Function Scale – Children and Adolescents Short Form

Multiple regression predicting UPT performance in the total sample. A multiple regression was performed to predict performance on the UPT (i.e., total items correct) from performance-based tasks and rating scales. A performance-based task composite was created using the two performance-based tasks. The model significantly predicted UPT performance, $F(2, 77)= 42.01$, $p < .001$, adj. $R^2= .51$. Both performance-based tasks and rating scales added significantly to the prediction $p < .001$. See Table 4.

Table 4

Multiple Linear Regression Predicting UPT Performance from Performance-Based Measures and Rating Scale with Total Sample

Variable	Standardized beta	*t*	Unique variance explained
Performance EF Composite	.64	8.02**	.40
BDEFS-CA	.24	2.94*	.05

*p <.01; **p <.001
Note: Performance EF Composite= Stroop-Colour Word test and Trail-Making test, BDEFS-CA= Barkley Deficits in Executive Function Scale – Children and Adolescents Short Form

Group Differences on Performance-Based Measures, Rating Scale, and UPT

Performance-based measures. Children with ADHD (mean rank= 34.50) showed lower performance on the TMT compared to TD children (mean rank= 45.93), U= 570, z= -2.20, p =.028. There were no significant differences between children with ADHD (mean rank= 36.08) and TD children's (mean rank= 44.50) performance on the Stroop test, U= 630, z= -1.62, p= .105.

Rating scale. EF as measured by the BDEFS was significantly lower in children with ADHD (mean rank= 21.03) compared to TD children (mean rank= 58.12), U=58, z= -7.13, p < .001.

Unstructured Performance-Based Task. In terms of the UPT scores children with ADHD (mean rank= 30.67) had significantly fewer correct items than TD children (mean rank 49.39), U= 424.50, z= -3.62, p < .001. Also, children with ADHD (mean rank= 45.57) circled significantly more items than TD children (mean rank= 35.92), U= 990.50, z= 2.32, p= .020. Additionally, children in the clinical group (mean rank= 50.64) left significantly more items blank than TD children (mean rank= 31.32), U= 1183.50, z= 3.83, p < .001. Children with ADHD (mean rank= 30.36) completed significantly fewer items compared to TD children (mean

rank= 49.68), *U*= 412.50, *z*= -3.83, *p* <.001. There were no significant differences in number of incorrect items between children in the clinical group (mean rank= 39.43) and TD children (mean rank= 41.46), *U*= 757.50, *z*= -.40, *p* = .691.

In regards to the UPT Totals children with ADHD (mean rank= 30.67) were significantly lower on the UPT Total Correct compared to TD children (mean rank= 49.39), *U*= 424.50, *z*= -3.62, *p* < .001. Children with ADHD (mean rank= 50.46) were significantly higher on the UPT Total Incorrect compared to TD children (mean rank= 31.49), *U*= 1176.50, *z*= 3.67, *p* < .001. Children with ADHD (mean rank= 30.51) were also significantly lower compared to TD children (mean rank= 49.54) on UPT Total Complete, *U*= 418.50, *z*= -3.77, *p* < .001. Children with ADHD (mean rank= 50.64) were significantly higher than TD children (mean rank=31.32) on UPT Total Incomplete, *U*= 1183.50, *z*= 3.83, *p* < .001. Finally, children in the clinical group (mean rank= 47.17) took significantly longer to complete the UPT compared to TD children (mean rank= 34.45), *U*= 1051.50, *z*= 2.44, *p*= .015.

As parent's marital status and mother's education were quite different between the TD and the ADHD group analyses were conducted in order to determine whether this had any effect on UPT performance. Marital status was dichotomized as married (i.e., married or common-law, *n*= 62) and unmarried (i.e., single, engaged, divorced, or separated, *n*= 17). Mother's education was dichotomized as low (i.e., up to Grade 8 and Grade 9-12, *n*= 6) and high (i.e., some post-secondary and completed post-secondary, *n*= 73). Significant differences in UPT performance were found. Children of parents who were married (mean rank= 43.14) had significantly higher UPT Total Items Correct compared to children who's parents were not married (mean rank= 28.56), U= 721.50, z= 2.33, *p*= .020. Children of parents who were married (mean rank= 36.86) had significantly lower UPT Total Items Incorrect compared to children who's parents were not

married (mean rank= 51.44), *U*= 332.50, *z*= -2.33, *p*= .020. Children of parents who were married (mean rank= 43.02) had significantly higher UPT Total Items Complete compared to children of parents who were not married (mean rank= 28.97), *U*= 714.50, *z*= 2.30, *p*= .021. Children of parents who were married (mean rank= 36.94) had significantly lower UPT Total Items Incomplete compared to children who's parents were not married (mean rank= 51.18), *U*= 337, *z*= -2.34, *p*= .020. Children of parents who were married (mean rank= 36.43) had significantly shorter completion times compared to children who's parents were not married (mean rank= 53.03), *U*= 305.50, *z*= -2.64, *p*= .008.

Significant differences also emerged in terms of mother's education. Children of mothers in the high education group (mean rank= 40.04) had significantly more UPT Total Items Correct compared to children of mother's in the low education group (mean rank= 16.60), *U*= 287, *z*= 2.31, *p*= .019. Children of mothers high education group (mean rank= 37.06) had significantly fewer UPT Total Items Incorrect compared to children of mothers in the low education group (mean rank= 59), *U*= 75, *z*= -2.16, *p* = .029. Children of mothers in the high education group (mean rank= 39.88) had significantly higher UPT Total Items Complete compared to children of mothers in the low education group (mean rank= 18.90), *U*= 275.50, *z*= 2.12, *p*= .037. Children of mothers in the high education group (mean rank= 37.37) had significantly lower UPT Total Items Incomplete compared to children of mothers in the low education group (mean rank= 56), *U*= 90, *z*= -1.89, *p*= .067. Finally, children of mothers in the high education group (mean rank= 36.93) had significantly shorter UPT Completion time compared to children of mothers in the low education group (mean rank= 60.80), *U*= 66, *z*= -2.34, *p*= .017.

It is important to note that sample sizes were grossly uneven for these analyses. Children with ADHD were disproportionately represented in the unmarried group and the low mother's

education group. Nevertheless, the significant results point to the importance of controlling for parent's marital status and mother's education in future studies.

In addition, an analysis was conducted to determine whether comorbid disruptive behaviour disorders (i.e., ODD or CD) had an effect on UPT performance. As only the ADHD group had positive ODD or CD diagnoses this analysis was conducted within the ADHD group. Diagnoses were dichotomized as ADHD only (n= 16) or ADHD+ODD and/or CD (n= 22). There were no significant differences on UPT Total Correct, UPT Total Incorrect, UPT Total Complete, UPT Total Incomplete, or UPT Completion time between children with ADHD only and children with ADHD and comorbid disruptive behaviour disorder, (p= .510; p= .510; p= .529; p= .529; p= .872, respectively).

UPT Behavioural Checklist. Children with ADHD (mean rank= 48.79) showed significantly more fidgety behaviours than TD children (mean rank= 33), U= 1113, z= -3.71, p < .001. Children with ADHD (mean rank= 44.92) showed significantly more off-task behaviours than TD children (mean rank= 36.50), U= 966, z= -3.11, p = .002. Children with ADHD (mean rank= 48.61) made significantly more verbalizations than TD children (mean rank= 33.17), U= 1106, z= 3.81, p < .001. Children with ADHD (mean rank= 43.26) showed significantly more verbal frustrations with the task than TD children (mean rank= 38), U= 903, z= 2.41, p= .016. Children with ADHD (mean rank= 41.61) did not show more physical frustration with the task than TD children (mean rank= 39.50), U= 840, z= 1.50, p= .135. Children with ADHD (mean rank 41.61) did not differ from TD children (mean rank= 39.50) in verbal refusal to do task, U= 840,

z= 1.50, p= .135. Children with ADHD (mean rank= 41.61) did not differ from TD children (mean rank= 39.50) in not physically engaging in the task, U= 840, z= 1.50, p= .135.

An analysis was conducted to determine whether comorbid disruptive behaviour disorders (i.e., ODD or CD) had an effect on the behaviours observed in the Behavioural Checklist. As only the ADHD group had positive ODD or CD diagnoses this analysis was conducted within the ADHD group. Diagnoses were dichotomized as ADHD only (n= 16) or ADHD+ODD and/or CD (n= 22). There were no significant differences in fidgety behaviours, off-task behaviours, physical frustration, verbal frustration, verbalizations, verbal refusal to do task, and not physically engaging in the task between children with ADHD only and children with ADHD and comorbid disruptive behaviour disorder (p= .312; p= .492; p= .942; p= .298; p= .510; p= .942; p= .942, respectively).

Binomial Logistic Regression Analyses Predicting Group Membership

Three logistic regressions were performed to determine whether UPT, performance-based measures, or the rating scale predicted group status (See Table 5). A performance-based measure composite was created using the two performance-based measures. The first model examined whether the UPT was a significant predictor of group status. The first model was statistically significant. Prediction success was 90.5% in the TD group and 52.6% in the ADHD group for an overall success rate of 72.5%. The second model examined whether the UPT and performance-based measures were significant predictors of group status. The second model was statistically significant, both the UPT and performance-based measures entered as significant predictors. Prediction success was 92.9% in the TD group and 55.3% in the ADHD group for an overall success rate of 75%. The third model examined whether the UPT, performance-based measures, and a rating scale were predictors of group status. The third model was statistically significant. Of the three predictor variables only the BDEFS-CA entered as a significant predictor. The

model correctly classified 90% of cases. Prediction success was 89.5% in the ADHD group and 90.5% in the TD group.

Table 5

Binomial Logistic Regression Predicting Group Membership with Total Sample

	B (SE)	Wald	Odds Ratio (95% CI)
Regression 1			
UPT Total Items Correct	-.19 (.06)	9.23**	.82 (.73, .93)
Cox & Snell R^2= .20			
Nagelkerke R^2= .27			
$\chi^2(1)$= 17.70***			
Regression 2			
UPT Total Items Correct	-.18 (.08)	5.93*	.83 (.72, .97)
Performance EF Composite	-.12 (.44)	.08	.89 (.38, 2.10)
Cox & Snell R^2= .20			
Nagelkerke R^2= .27			
$\chi^2(2)$ = 17.78***			
Regression 3			
UPT Total Items	.01 (.07)	.006	1.01 (.87, 1.16)

Correct			
Performance EF	-1.76 (.91)	3.77	.17 (.03, 1.02)
Composite			
BDEFS-CA	-.28 (.07)	16.05***	.76 (.70, .87)
Cox & Snell R^2= .62			
Nagelkerke R^2= .82			
$\chi^2(3)$= 76.59***			

*p <.05; ** p <.01; *** p < .001
Note: Performance EF Composite= Stroop-Colour Word test and Trail-Making test, BDEFS-CA= Barkley Deficits in Executive Function Scale – Children and Adolescents Short Form

Discussion

The purpose of this study was to examine the role of structure in performance-based measures as an explanation for the discordance between EF rating scales and performance-based tasks. In order to do this a novel Unstructured Performance-Based Task (UPT) was developed. The aim in designing this task was to minimize the structure imposed by the examiner and to leave interpretation of the task somewhat open to the examinee. This study had three objectives. First, to examine the pattern of associations between performance-based measures, rating scale, and the UPT as well as the associations between the UPT, age, and intellectual ability. Second, to compare EF as measured by rating scales, performance-based measures, and the UPT in children with ADHD and TD children. Third, to determine whether the UPT significantly predicted ADHD status. Results showed that there were significant associations among the UPT, performance-based measures, rating scales, age and intellectual ability. Multiple regression analyses showed that performance-based measures and rating scales significantly predicted UPT performance. Additionally, there were group differences between children with ADHD and TD

children on the UPT and the accompanying Behavioural Checklist. Finally, logistic regression analyses showed that the UPT predicted ADHD status above and beyond performance-based measures, however when rating scales were entered as a predictor the UPT was no longer significant.

Associations Among the UPT, Performance-Based Measures, and Rating Scale

Several significant associations emerged among the UPT variables, performance-based measures, and rating scales.. First, consistent with past literature, associations between both of the performance-based measures and the rating scale were small to moderate and not significant (Bodnar et al., 2007; Gray et al., 2015; Mahone et al., 2002; McAuley et al., 2010; Toplak et al., 2013). However, performance on the UPT, specifically UPT Total Items Correct, UPT Total Items Incorrect, UPT Total Items Completed, and UPT Completion time were associated with the BDEFS-CA, the Stroop test, and the TMT. These patterns of associations among the UPT variables, performance-based measures, and rating scales suggests the UPT may be tapping into EF as measured by performance-based measures and rating scales. The strength of the correlations ranged from small to moderate with the smallest correlation being between UPT Total Complete and the Stroop test (.23) and the strongest correlation being between the UPT Total Correct and TMT (.55), correlations between the UPT and the BDEFS-CA were small to moderate (.24-.37). These associations support that the UPT is assessing the underlying construct of EF. Furthermore, it is notable that there were no significant associations between the performance-based measures and rating scales, but both of these measures were significantly associated with the UPT. This finding suggests the UPT may be assessing a novel or additional aspect of EF above and beyond what performance-based tasks and rating scales measure on their own.

Multiple regression predicting UPT performance in the total sample. A multiple regression was conducted to determine the contributions of performance-based measures and rating scales in predicting performance on the UPT. Results showed that both the performance-based measure composite and rating scale were significant predictors of UPT performance, as measured by UPT Total Correct Items, This finding provides further support that the UPT is assessing the underlying construct of EF. Although both measures entered as significant predictors, performance-based measures accounted for more variance than rating scales (40% versus 5%). This finding is not surprising as the UPT is assessing children's performance in a similar manner to performance-based measures (i.e., tasks are completed by the child one to one testing). The important difference is that the UPT imposes less structure on the child, and allows for more interpretation of task demands left open to the examinee compared to performance-based tasks. However, it is notable that the rating scale still entered as a significant predictor of UPT performance. The finding that both measures entered as predictors provides further support that the UPT may be tapping into an additional or novel aspect of EF that is not be assessed by either rating scales or performance-based measures alone.

Associations Among the UPT, Age, and Intellectual Ability

The UPT was significantly associated with both intellectual ability and age, which is what was expected as both of these have been shown to be associated with EF (Arffa, 2007; Brydges, Reid, Fox, & Anderson, 2012; Romine & Reynolds, 2005; Miyake & Friedman, 2012; Zelazo et al., 2004).

Group Differences on Performance-Based Measures, Rating Scales, and the UPT

Group differences were found between children with ADHD and TD children on performance-based measures, rating scales, and the UPT. First, consistent with past literature,

children with ADHD had higher levels of executive dysfunction as indicated by the BDEFS-CA and the TMT compared to TD children. However, no significant differences were found between children with ADHD and TD children on the Stroop task. This lack of difference may be due to the different cognitive abilities assessed by each task (i.e., set-shifting versus cognitive inhibition) (Nigg et al., 2005). It is also possible the Stroop was not sensitive enough to detect EF deficits in our sample. Group differences emerged on all aspects of UPT performance, children in the ADHD group completed fewer items, had fewer correct items, and took longer to complete the task. These findings support that the UPT is capturing differences between children with ADHD and TD children. It is important to note that when comparing children with ADHD and children with ADHD+ODD and/or CD there were no differences attributable to comorbid disorders. This finding reinforces that the UPT is capturing differences related to ADHD. Finally, group differences between children with ADHD and TD children were also found on the UPT Behavioural Checklist. Specifically, children with ADHD were rated as showing significantly greater frequencies of fidgety (e.g., squirming in chair, tapping table with pencil) and off-task behaviours (e.g., doodling on sheet, playing with pencil) as well as made more verbalizations (e.g., singing, humming), and exhibited greater verbal frustration with the task (e.g., muttering, sighing, grunting). Past research has shown that when undergoing one-to-one testing children with hyperactive behaviours may present as calm and attentive, especially when explicit directions are given and task duration is short (Draeger et al., 1986; Mahone & Hoffman, 2007). It is possible that by minimizing the structure in this task, we were able to access how a children may behave in a more unstructured setting (e.g., the classroom, free play). Again, no differences were found between children with ADHD and children with ADHD+ODD and/or

CD on the Behavioural Checklist, providing support that the behaviours noted were attributable to ADHD and not comorbid disorders.

The Role of Performance-based Measures, Rating Scales, and the UPT in Predicting ADHD Diagnosis

Binomial logistic regression was used to determine the contributions of the performance-based measure composite, rating scale, and the UPT in predicting ADHD diagnosis. When entered as the sole predictor the UPT significantly predicted ADHD diagnosis. Sensitivity was 52.6% and specificity was 90.5%. When the performance-based task composite was added both the UPT and performance-based measures significantly predicted ADHD diagnosis. There was a slight increase in successful prediction, sensitivity was 55.3% and specificity was 92.9%. However, when the rating scale was added only the rating emerged as a significant predictor of ADHD diagnosis. Sensitivity was 89.5% and specificity was 90.5%. These results suggest that the UPT is able to differentiate between children with ADHD and TD children, but not over and above the rating scale. This finding may be due to common variance shared between rating scales and the UPT. Furthermore, rating scales ask questions that are closely related to ADHD symptoms (McAuley et al., 2010). For this reason it is unsurprising that rating scales were more predictive of ADHD diagnosis compared to performance-based tasks and the UPT, which were designed to assess EF, not to diagnose ADHD.

Implications

The patterns of associations, group differences, and regression analyses all support the UPT as a promising measure to assess EF in children with ADHD. The UPT was significantly and moderately associated with both performance-based measures and rating scales of EF, whereas performance-based tasks and rating scales were not significantly associated. This

suggests that the UPT is assessing the underlying construct of EF, and that it is measuring an aspect of EF that is not assessed by either performance-based tasks or rating scales alone.

Group differences between children with ADHD and TD children also provide support for the UPT assessing EF. As children with ADHD have been shown in the literature to have EF deficits, it is expected that group differences would emerge on a measure assessing this construct (Barkley & Fischer, 2011; Barkley & Murphy, 2010; Biederman et al., 2007; Mahone et al., 2002; Mahone & Hoffman, 2007; Toplak et al., 2005; Nigg et al., 2005; Scheres et al., 2004).

Furthermore, minimizing the structure in a performance-based task allowed for observation of children in typical performance situations rather than optimal performance situations. Recall that typical performance situations are those in which no overt instructions to maximize performance are given and task interpretation is primarily left up to the examinee (Toplak et al., 2013). Whereas, in optimal performance situations task interpretation is highly constrained and there are clear instructions given as to how to maximize performance (Toplak et al., 2013). Traditional standardized performance-based tasks occur in optimal conditions. Clear task instructions on what is considered maximal performance are given to the examinee. This approach greatly alleviates the self-regulation demanded of the examinee (Clark et al., 2000). Thus, the circumstances during performance-based tasks are not necessarily reflective of the manner in which an examinee would approach tasks that require EF in everyday activities where there are numerous demands and self-regulation is required. By minimizing structure imposed by the examiner, and not giving explicit instructions on how to maximize performance the UPT allowed for observation of children in a typical performance situation, where self-regulation is demanded of them.

Current assessment practices are to include both performance-based tasks of EF and rating scales (McAuley et al., 2010; Toplak et al., 2013). However, these measures are both limited in that performance-based tasks are highly structured and rating scales are based on retrospective reporting of behaviours related to EF in everyday activities. Both of these techniques have strengths and provide valuable, non-redundant information about the child (Toplak et al., 2013). However, what is currently lacking is a measure of how a child would perform in an unstructured situation. This information is very valuable as children are faced with many unstructured situations in which self-regulation is important (i.e., free play, independent class work). Furthermore, past research has shown children with ADHD can present as attentive and calm in one-to-one testing when instructions are explicit and tasks are simple and short (Mahone & Hoffman, 2007). This may be leading to "missing" the EF deficits children with ADHD show in their everyday activities. By including an unstructured measure, such as the UPT, in an assessment one would be able to learn information about how the child performs in less structured environments where self-regulation is demanded of them.

Future Directions

Future research should be undertaken in order to further investigate the role of structure (or lack thereof) in the UPT. One way of doing so is by designing a parallel instrument with a similar style of questions that is presented in the exact same way (i.e., 11x17 piece of paper) but with questions being presented in a structured grid, rather than in a scattered, random order. The original UPT and this structured version would be administered to the same participants in a within-subjects design. Comparing performance on the UPT and the parallel structured form allows for further elucidation of the role of structure in performance-based tasks.

Limitations

There are some important limitations to note. Firstly, mother's age and parent's marital status were quite different between the two groups. Analyses did show group differences on the UPT between children of married and unmarried parents, and children of mother's with high education and low education. However, children with ADHD were disproportionately represented in the unmarried parents and low mother's education group, which could explain these findings. Future studies with larger sample sizes could test for any significant associations between EF as measured by the UPT and these variables. Additionally, examiners were not masked to ADHD diagnosis. This could have potentially influenced them when completing the UPT Behavioural Checklist. Future studies should mask examiners to any diagnoses in order to avoid potential bias. Finally, future studies should also inquire about any neurological conditions (e.g., traumatic brain injury).

Conclusion

The current study examined a novel Unstructured Performance-Based Task (UPT) in children with ADHD and TD children. Significant and moderate associations were found between the UPT and performance-based tasks as well as the rating scale. Additionally, there were no significant associations between the rating scale and either performance-based task. Furthermore, performance-based tasks and the rating scale significantly predicted UPT performance. Significant associations between the UPT and age and intellectual ability were also found. Taken together, these findings provide support that the UPT is assessing EF above and beyond what performance-based tasks and rating scales measure individually. Notably, group differences emerged on the UPT between children with ADHD and TD children. Children with ADHD completed fewer items, had fewer correct answers, and took longer to complete the UPT

compared to TD children. Furthermore, children with ADHD displayed more behaviours from the Behavioural Checklist compared to TD children. Using logistic regression analyses the UPT significantly predicted ADHD status. When a performance-based task composite was entered the UPT remained as a significant predictor. However, when the rating scale was added neither the performance-based composite nor the UPT were significant predictors of ADHD status. This could be because performance-based tasks and the UPT were designed to assess EF not diagnose ADHD, whereas rating scales ask questions that are closely related to ADHD symptoms. Results indicate the UPT may be a promising measure to assess EF related difficulties in ADHD and provide a picture of children's behaviours in unstructured environments.

References

Achenbach, T. M., & Rescorla, L. A. (2001). The manual for the ASEBA school-age forms & profiles. Burlington, VT: University of Vermont, Research Center for Children, Youth, and Families.s

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (5th ed.).* Washington, DC: Author.

Arffa, S. (2007). The relationship of intelligence to executive function and non-executive function measures in a sample of average, above average, and gifted youth. *Archives of Clinical Neuropsychology*, *22*(8), 969-978.

Barkley, R. A. (2012). *Barkley Deficits in Executive Functioning Scale--Children and Adolescents (BDEFS-CA).* Guilford Press.

Barkley, R. A., & Fischer, M. (2011). Predicting impairment in major life activities and occupational functioning in hyperactive children as adults: Self-reported executive function (EF) deficits versus EF tests. *Developmental Neuropsychology*, 36(2), 137-161.

Barkley, R. A., & Murphy, K. R. (2010). Impairment in occupational functioning and adult ADHD: the predictive utility of executive function (EF) ratings versus EF tests. *Archives of Clinical Neuropsychology, 25*(3), 157-173.

Biederman, J. (2005). Attention-deficit/hyperactivity disorder: a selective overview. *Biological Psychiatry*, *57*(11), 1215-1220.

Biederman, J., Monuteaux, M. C., Doyle, A. E., Seidman, L. J., Wilens, T. E., Ferrero, F., ... & Faraone, S. V. (2004). Impact of executive function deficits and attention-deficit/hyperactivity disorder (ADHD) on academic outcomes in children. Journal of consulting and clinical psychology, 72(5), 757.

Biederman, J., Petty, C. R., Fried, R., Black, S., Faneuil, A., Doyle, A. E., ... & Faraone, S. V. (2007). Discordance between psychometric testing and questionnaire-based definitions of executive function deficits in individuals with ADHD. *Journal of Attention Disorders, 12*(1), 92-102.

Bodnar, L. E., Prahme, M. C., Cutting, L. E., Denckla, M. B., & Mahone, E. M. (2007). Construct validity of parent ratings of inhibitory control. *Child Neuropsychology*, *13*(4), 345-362.

Brydges, C. R., Fox, A. M., Reid, C. L., & Anderson, M. (2014). The differentiation of executive functions in middle and late childhood: A longitudinal latent-variable analysis. *Intelligence*, *47*, 34-43.

Carlson, S. M., Zelazo, P. D., & Faja, S. (2013). Executive function. In P. D. Zelazo (Ed.), Handbook of developmental psychology. New York, NY: Oxford University Press.

Chan, R. C., Shum, D., Toulopoulou, T., & Chen, E. Y. (2008). Assessment of executive functions: Review of instruments and identification of critical issues. *Archives of Clinical Neuropsychology*, *23*(2), 201-216.

Clark, C., Prior, M., & Kinsella, G. J. (2000). Do executive function deficits differentiate between adolescents with ADHD and oppositional defiant/conduct disorder? A neuropsychological study using the Six Elements Test and Hayling Sentence Completion Test. *Journal of abnormal child psychology*, *28*(5), 403-414.

Delis, D.C., Kaplan, E., & Kramer, J.H. (2001). Delis-Kaplan Executive Function System examiner's manual. San Antonio, Texas: NCS Pearson, Inc

Draeger, S., Prior, M., & Sanson, A. (1986). Visual and auditory attention performance in

hyperactive children: Competence or compliance. *Journal of Abnormal Child Psychology, 14*(3), 411-424.

Ek, U., Westerlund, J., & Fernell, E. (2013). General versus executive cognitive ability in pupils with ADHD and with milder attention problems. *Neuropsychiatric Disease and Treatment, 9,* 163-168.

Emslie, H., Wilson, F.C., Burden, V., Nimmo-Smith, I., & Wilson, B.A. (2003). Behavioural assessment of the dysexecutive dyndrome for children (BADS-C). London, England: Harcourt Assessment.

Evans, A. S., & Preston, A. S. (2011). Test of Everyday Attention for Children. In *Encyclopedia of Clinical Neuropsychology* (pp. 2493-2493). Springer New York.

Fisher, L., Lucas, C., Sarsfield, & Shaffer (2006). *Interviewer manual*. Columbia University DISC Development Group.

Gioia, G. A., Isquith, P.K., & Guy, S.C., & Kentworthy, L. (2000) *Behaviour rating inventory of executive function.* Florida: Psychological Assessment Resources

Gregory, R.J. (2011). *Psychological testing history, principles, and applications (6th Ed.).* Boston: Pearson, Inc.

Golden, C. J. (1978). Stroop colour and word test. *Age*, *15*, 90.

Gray, S. A., Fettes, P., Woltering, S., Mawjee, K., & Tannock, R. (2015). Symptom Manifestation and Impairments in College Students With ADHD. *Journal of learning disabilities*, 1-15.

Heaton, R. K. chelune, GJ, Talley, JL, Kay, GG, & curtis, G.(1993). *Wisconsin Card Sorting Test manual: Revised and expanded*. Florida: Psychological Assessment Resources

Holtz, J.L. (2011). *Applied clinical neuropsychology an introduction.* New York: Springer

Publishing Company.

Imeraj, L., Antrop, I., Roeyers, H., Deboutte, D., Deschepper, E., Bal, S., & Sonuga-Barke, E. (2016). The Impact of Idle Time in the Classroom Differential Effects on Children With ADHD. *Journal of Attention Disorders*,*20*(1), 71-81.

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman brief intelligence test*. John Wiley & Sons, Inc.

Lawrence, V., Houghton, S., Tannock, R., Douglas, G., Durkin, K., & Whiting, K. (2002). ADHD outside the laboratory: Boys' executive function performance on tasks in videogame play and on a visit to the zoo. *Journal of Abnormal Child Psychology*, *30*(5), 447-462.

Lawrence, V., Houghton, S., Douglas, G., Durkin, K., Whiting, K., & Tannock, R. (2004). Executive function and ADHD: A comparison of children's performance during neuropsychological testing and real-world activities.*Journal of attention disorders*, *7*(3), 137-149.

Lezak, M. (1995). Neuropsychological testing. *Oxford*: University Press

Loo, S.K., Humphrey, L.A., Tapio, T., Moilanen, I.K., McGough, J.J., McCracken,J.T., …Smalley, S.L. (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., & Hoffman, J. (2007). Behavior ratings of executive function among preschoolers with ADHD. *The Clinical Neuropsychologist*, *21*(4), 569-586.

Mahone, E. M., Cirino, P. T., Cutting, L. E., Cerrone, P. M., Hagelthorn, K. M., Hiemenz, J. R., ... & Denckla, M. B. (2002). Validity of the behavior rating inventory of executive

function in children with ADHD and/or Tourette syndrome. Archives of Clinical Neuropsychology, 17(7), 643-662.

McAuley, T., Chen, S., Goos, L., Schachar, R., & Crosbie, J. (2010). Is the behavior rating inventory of executive function more strongly associated with measures of impairment or executive function?. *Journal of the International Neuropsychological Society*, *16*(03), 495-505.

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions four general conclusions. *Current Directions in Psychological Science, 21*(1), 8-14.

Nigg, J.T., Willcutt, E.G., Doyle, A.E., & Sonuga-Barke, E.J.S. (2005). Causal heterogeneity in attention-deficit/hyperactivity disorder: do we need neuropsychological impaired subtypes? *Biological Psychiatry, 57(11),* 1224-1230.

Osterrieth, P. A. (1944). Le test de copie d'une figure complexe [A test of copying a complex figure]. *Archives de Psychologie*, 30, 206-356.

Pennington, B. & Ozonoff, S. (1996). Executive functions and developmental psychopathology. *Journal of Child Psychology and Psychiatry, 37*(1), 51-87.

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. *The American Journal of Psychiatry, 164*(6), 942-948.

Qian, Y., Shuai, L., Cao, Q., Chan, R. C., & Wang, Y. (2010). Do executive function deficits differentiate between children with attention deficit hyperactivity disorder

(ADHD) and ADHD comorbid with oppositional defiant disorder? A cross-cultural study using performance-based tests and the behavior rating inventory of executive function. *The Clinical Neuropsychologist*, *24*(5), 793-810.

Reitan, R. M. (1971). Trail Making Test results for normal and brain-damaged children. *Perceptual and Motor Skills*, 33, 575-58.

Romine, C. B., & Reynolds, C. R. (2005). A model of the development of frontal lobe functioning: Findings from a meta-analysis. *Applied Neuropsychology, 12*(4), 190-201.

Roth, R. M., Isquith, P. K., & Gioia, G. A. (2005). Behavior rating inventory of executive function-adult version (BRIEF-A). *Lutz, FL: Psychological Assessment Resources*.

Salthouse, T. A., Atkinson, T. M., & Berish, D. E. (2003). Executive functioning as a potential mediator of age-related cognitive decline in normal adults. *Journal of Experimental Psychology: General*, *132*(4), 566.

Scheres, A., Oosterlaan, J., Geurts, H., Morein-Zamir, S., Meiran, N., Schut, H., ... & Sergeant, J. A. (2004). Executive functioning in boys with ADHD: primarily an inhibition deficit? *Archives of Clinical Neuropsychology, 19*(4), 569-594.

Stanovich, K. E. (2009). Distinguishing the reflective, algorithmic, and autonomous minds: Is it time for a tri-process theory. *In two minds: Dual processes and beyond*, 55-88.

Strauss, E., Sherman, E. M., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary*. New York: Oxford University Press.

Thorell, L.B., Eninger, L., Brocki, K.C., & Bohlin, G. (2010). Childhood Executive Function Inventory (CHEXI): A promising measure for identifying young children with ADHD? *Journal of Clinical and Experimental Neuropsychology, 32,* 536-552.

Toplak, M. E., Bucciarelli, S. M., Jain, U., & Tannock, R. (2008). Executive functions: performance-based measures and the behavior rating inventory of executive function (BRIEF) in adolescents with attention deficit/hyperactivity disorder (ADHD). *Child Neuropsychology, 15*(1), 53-72.

Toplak, M. E., West, R. F., & Stanovich, K. E. (2013). Practitioner Review: Do performance‐based measures and ratings of executive function assess the same construct? *Journal of Child Psychology and Psychiatry*, *54*(2), 131-143.

Wechsler, D. (2014). *Wechsler Intelligence Scale for Children- Fifth Edition.* San Antonio, TX: Psychological Corporation.

Welsh, M. C., & Pennington, B. F. (1988). Assessing frontal lobe functioning in children: Views from developmental psychology. *Developmental Neuropsychology*, *4*(3), 199-230.

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.

Wilson, B. A., Evans, J. J., Alderman, N., Burgess, P. W., & Emslie, H. (1997). Behavioural assessment of the dysexecutive syndrome. *Methodology of frontal and executive function*, 239-250.

Zelazo, P. D., Craik, F. I., & Booth, L. (2004). Executive function across the life span. *Acta Psychologica, 115*(2), 167-183.

Appendix A

Unstructured Performance Measures used to Assess Executive Functioning

Measure	**Description**	**Unstructured Components**	**Structured Components**
Design fluency test (Delis Kaplan Executive Function System; Delis, Kaplan, & Kramer, 2001)	This task measures how many different designs the examinee can draw in 60 seconds. The examinee is presented with rows of boxes containing arrays of dots and asked to draw a different design in each box. There are four conditions with specified rules pertaining to how the dots may be joined.	• The examinee can create any designs they want as long as they conform to the specified rules of each condition	• Rules of each condition, limits the amount of designs the examinee can make. • Time limit.
Zoo Map Tests 1 & 2 (BADS-C; Emslie et al., 2003)	In these tasks the child is asked to plan a route in order to visit 6 out of 12 possible locations at a zoo while accounting for specified restrictions, and starting and stopping points.	• Zoo Map Test 1 is an open-ended task where little structure is provided by the examiner. In this task the child must work out a plan in advance in order to minimize errors.	• In Zoo Map Test 2 structure is imposed and the child is told to show what animals they would like to visit but they must visit each animal in the order they are presented on the map. Thus, in Zoo Map Test 2 the demand of planning on the child is much lower than in the unstructured Zoo Map Test 1.

The Rey-Osterrieth Complex Figure Test (ROCF; Osterrieth, 1944)	This task is measure of visuoconstructional ability and visual memory. It has also been found to provide a measure of EF, in particular, planning and organization (Watanbe et al., 2005). In this task the examinee is asked to reproduce a complex figure drawing presented to them. There are three versions, copy, immediate recall, and delayed recall. One method of scoring this task is the Boston Qualitative Scoring System (Stern et al., 1999). According to this system all components of the figure belong to one of three groups: configural elements, clusters, and details. Furthermore, these groups are hierarchal with respect to structural importance with configural elements being at the top of the hierarchy and details being at the bottom.	• The examinee can approach drawing the figure in multiple ways. For example, Examinees may choose to take a more global approach to copying and or memorizing and begin drawing the figure with the configural elements first, then adding clusters, and finally adding details. Alternatively, examinees may focus on more focal aspects of the figure such as clusters and details while neglecting the global figure as a whole.	• The same complex design is presented to each examinee to replicate. • The scoring system specifies what groups elements of the figure belong to.

Map Mission (The Test of Everyday Attention for Children [TEA-Ch] Evans & Preston, 2011)	Map mission is a selective attention task. In Map Mission the examinee must locate 80 targets, which are small knife-and-fork restaurant symbols that are randomly distributed across a map interspersed with distracter items of similar size. The examinee is asked to circle all of the target items in 1 minute. It is actually unlikely that the examinee will be able to circle all of the targets in 1 minute and the examinee is scored based on how many targets they marked correctly.	• The random distribution of items on the map	• Time limit.
Cancellation (Wechsler Intelligence Scale for Children – Fifth Edition [WISC-V] Wechsler, 2014)	Cancellation was designed to measure processing speed. In this task the examinee is asked to scan an arrangement of pictures and mark target pictures within a specified time limit of 45 seconds. There are two parts to this task, in one the arrangement of pictures is in a structured order, in the second the arrangement of pictures are in a random order. The examinee is scored based on items marked correctly (targets) and items marked incorrectly (non-targets).	• In the second component of the task the arrangement of the pictures is in random order.	• Time limit.

Appendix B

Unstructured Performance-Based Task and Instructions

4+0=

3X2=

Name a colour.

9-4=

Write your name:

Give a word that ends with the letter G.

What is this a picture of?

Do pickle and bickle rhyme?

Draw a tree.

Name a zoo animal.

5+4=

5-2=

What rhymes with face?

How many squares?

3X4=

What is the opposite of small?

How many legs does a duck have?

9X3=

Copy this pattern:
XOXXOO

3+7=

Is this a triangle?

What rhymes with hat?

Is pizza a food?

Give a word that has 6 letters.

7-1=

Do feet and meat rhyme?

How many letter t's are in this sentence:
This turtle ate tulips.

Put a dot in each circle.

Give a word that starts with the letter M.

Write down your birthday.

8+4=

Draw a circle.

Is 7 bigger than 3?

Write a number.

Do airplanes have wings?

8-3=

5X2=

Do bark and part rhyme?

Write a word.

Finish the sentence:

Birds live in a

Name something bigger than an ant.

Is zunk a word?

“I would like you to complete the following worksheet. If you do not know the answer for any of the problems, just circle it and go on to the next problem. I cannot read any of the questions to you. Just do your very best, and when you are done, please bring the worksheet to me.”

This is meant to be an unstructured task. The examiner should sit in another part of the room while the child is working on this task.

If the child displays oppositional behaviour (verbally or physically refusing to do the task e.g., I don’t want to do this, or sitting with arms crossed and not picking up the pencil) the examiner may give **ONE** prompt:

“Please complete the worksheet. Remember you may circle items if you do not know the answer”

Note that the prompt was given and what occurred after the prompt (e.g., child continued to refuse to do task, child did task immediately after prompt, child refused to do task for a while but proceeded to task eventually or any thing else that happened) on the behavioural observations sheet.

Appendix C

Scoring template for the Unstructured Performance-Based Task

ADHD Performance Task – Answer Key

Using a red pen:
Checkmark correct items directly on the task.
Underline incorrect items directly on the task.

Item	Correct (0)	Incorrect (1)	Circled (2)	Blank (3)
1. *4+0* = 4				
2. *Give a word that ends with the letter G.* Any word in English that ends with G. Do not penalize for spelling if it's easily recognizable what the child meant (e.g., spelling instead of spelling is acceptable). Can be a name (e.g., Greg).				
3. *3x2*= 6				
4. *Name a colour.* Any colour. Child can attach quantifier to colours, e.g., cherry red, light blue. Do not penalize for spelling if it's easily recognizable what the child meant.				
5. *9-4* = 5				
6. *What is this a picture of?* House or home				
7. *Write your name:* Check name in Participant ID Excel document. First name only and first and last name are both correct. Do not penalize for incorrect capitalization.				
8. *Draw a tree.* Any picture resembling a tree gets 1 point.				
9. *Name a zoo animal.* Any animal found in a zoo, e.g., zebra, giraffe, monkey etc.				
10. *Do pickle and bickle rhyme?* Yes.				
11. *9x3*= 27				
12. *5-2*= 3				
13. *How many legs does a duck have?* Two or 2.				
14. *3x4*= 12				
15. *Copy this Pattern:* XOXXOO				

16. *What is the opposite of small?* Big or large.				
17. *3+7*= 10				
18. *What rhymes with face?* Any word in English that rhymes with face, e.g., place, space, mace. Do not penalize for spelling if it's easily recognizable what the child meant.				
19. *5+4*= 9				
20. *How many squares?* 7 or seven.				
21. *Do feet and meat rhyme?* Yes.				
22. *Is this a triangle?* Yes.				
23. *What rhymes with hat?* Any word in English that rhymes with hat, e.g., bat, mat, cat. Do not penalize for spelling if it's easily recognizable what that child meant.				
24. *How many letter t's are in this sentence: This turtle ate tulips.*5 or five.				
25. *Is pizza a food?* Yes.				
26. Give a word that has 6 letters. Any word in English that is 6 letters long. Can be a name. Do not penalize for spelling if it's easily recognizable what the child meant unless it interferes with the prompt (i.e., if the spelling error makes a word shorter or longer than it really is, e.g., magic is five letters but misspelled as magick it's 6 letters this would **NOT** get a point because the spelling error interfered with the letter count).				
27. *7-1*= 6				
28. *Put a dot in each circle.* There should be a dot in each of the 6 circles. Do not penalize for messy dot placement as long as it's inside the circle and the mark is identifiable as a dot.				
29. *Give a word that starts with the letter M.* Any word in English that starts with M, can be a name or a place. Do not penalize for spelling errors if it is easily recognizable what the child meant.				
30. *Write down your birthday.* Can be in any format (e.g., month can be written as October, Oct, 8) and can be in any order (month/day/year, day/month/year,				

year/month/day, year/day/month). Confirm with date on K-BIT 2.				
31. *8+4*= 12				
32. *Draw a circle.* Child gets a point for drawing a circle. Do not penalize for messiness if the shape is easily recognizable as a circle.				
33. *8-3*= 5				
34. *Is 7 bigger than 3?*Yes.				
35. *5x2*= 10				
36. *Write a number.* Any number written in digit(s).				
37. *Do airplanes have wings?* Yes.				
38. *Is zunk a word?* No.				
39. *Do bark and part rhyme?* No.				
40. Write a word. Any word in English, can be a place or a name. Do not penalize for spelling if it's easily recognizable what the child meant.				
41. *Finish this sentence: Birds live in a* Nest				
42. Name something bigger than an ant. Anything bigger than an ant, can be an object, place, or person. Do not penalize for spelling if it's easily recognizable what the child meant				
Totals				
Total completed (sum of three above)				

Appendix D

Accompanying Behavioural Checklist to the Unstructured Performance-Based Task

Experimenter rated behaviours during Unstructured Performance Task
1=not at all; 2=a little bit; 3=very much

1. Fidgety behaviours (squirming in chair, tapping table with pencil)
   Rating: _____ Indicate Behaviour:
   _____
2. Off-task behaviours (doodling on sheet, playing with pencil)
   Rating: _____ Indicate Behaviour:
   _____
3. Expressed physical frustration with task (hitting table with fist)
   Rating: _____ Indicate Behaviour:
   _____
4. Verbal frustration with task (muttering, sighing, grunting)
   Rating: _____ Indicate Behaviour:
   _____
5. Verbalizations (singing, humming, just making noise to make noise)
   Rating: _____ Indicate Behaviour:
   _____

Oppositional Behaviours during Unstructured Performance Task:
1=not at all; 2=a little bit; 3=very much

1. Verbally refusing to do the task
   Rating: _____ Indicate Behaviour:
   _____

2. Not physically engaging in the task (sitting with arms crossed, not picking up pencil, looking away from task etc.)
   Rating: _____ Indicate Behaviour:
   _____

Behavioural observation, additional notes:

_____
_____
_____
_____

How did child approach task? Describe strategy for completing all of the items.

_____
_____
_____
_____

Total time taken to complete this task: _____
Based on oppositional behaviours is the task valid (Y/N)? _____