2014

The ecological validity of neuropsychological tests of executive function in children with Fetal Alcohol Spectrum Disorder (FASD)

Jaspreet Kaur Rai
University of Windsor

The Ecological Validity of Neuropsychological Tests of Executive Function in Children with Fetal Alcohol Spectrum Disorder (FASD)

By

Jaspreet Kaur Rai

A Thesis
Submitted to the Faculty of Graduate Studies
through the Department of Psychology
in Partial Fulfillment of the Requirements for
the Degree of Master of Arts
at the University of Windsor

Windsor, Ontario, Canada

2014

The Ecological Validity of Neuropsychological Tests of Executive Function in Children with Fetal Alcohol Spectrum Disorder (FASD)

by

Jaspreet Kaur Rai

APPROVED BY:

D. Kane
Faculty of Nursing

A. Baird
Department of Psychology

J. Casey, Advisor
Department of Psychology

August 19, 2014

DECLARATION OF ORIGINALITY

I hereby certify that I am the sole author of this thesis and that no part of this thesis has been published or submitted for publication.

I certify that, to the best of my knowledge, my thesis does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained a written permission from the copyright owner(s) to include such material(s) in my thesis and have included copies of such copyright clearances to my appendix.

I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee and the Graduate Studies office, and that this thesis has not been submitted for a higher degree to any other University or Institution.

ABSTRACT

Although performance-based tests are used to evaluate executive function (EF) processes, studies comparing scores from performance-based and behavioural measures of EF indicate that the former have little, if any, ecological validity in various clinical populations. This study examined the relationship between three performance-based EF tests—the Wisconsin Card Sorting Test (WCST), Trail Making Test (TMT), and Controlled Oral Word Association Test (COWAT)—and the Behavior Rating Inventory of Executive Function (BRIEF) in a predominantly Aboriginal/Indian sample of children with Fetal Alcohol Spectrum Disorder (FASD; $N = 96$). Bivariate correlations and canonical correlation analysis were not statistically significant, although more impaired scores on BRIEF Organization of Materials were weakly associated with fewer perseverative errors on the WCST. Performance-based test variables also did not meaningfully predict scores on the two BRIEF indexes. Findings are discussed in the context of the different aspects of EF assessed by performance-based EF tests and the BRIEF.

ACKNOWLEDGEMENTS

First and foremost, I extend my deepest appreciation to Dr. Lloyd Flaro (Child Clinical Neuropsychologist, Edmonton, Alberta, Canada) for his diligence and dedication in compiling an incredibly large research database over nearly 20 years of clinical work, and for being so kind to share that database with me.

I also express my sincerest gratitude to my supervisor, Dr. Joseph Casey, and my committee members, Dr. Anne Baird and Dr. Debbie Kane, for their insight, guidance, and feedback throughout the course of this project.

I especially thank Dr. Dennis Jackson, not only for his humour and non-judgment, which I have found to be ever so refreshing during my first two years of graduate school, but also for his constant willingness to share his statistical expertise—even at bizarre hours of the night. (Dr. Jackson once replied to one of my frantic e-mails about canonical correlation analysis at one o'clock in the morning).

I am eternally grateful for having had the opportunity to work with Dr. Paul Green (Clinical Neuropsychologist, Edmonton, Alberta, Canada), who is truly the epitome of a scientist-practitioner. Your knowledge, your passion, and your contribution to neuropsychology will never cease to inspire me.

Last, but certainly not least, I thank my family, without whose unwavering support (both financial and emotional) my professional dream would be just that: a dream.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Executive function (EF) is an umbrella term that is used to describe control or self-regulatory functions that organize and direct cognitive activity, emotional responses, and overt behaviour (Gioia & Isquith, 2004). Some theorists have differentiated between metacognitive and emotional/motivational EF (Fuster, 2001, 2002; Happaney, Zelazo, & Stuss, 2004), where metacognitive EF allows an individual to organize his behavioural response for the purpose of solving a problem or attaining an external goal, and emotional/motivational EF coordinates cognition and emotion/motivation so that an individual can fulfill his basic impulses in a socially acceptable manner (Cummings, 1993). While it seems reasonable that both domains of EF would be important to an individual's functioning in the real world, traditional neuropsychological assessment has focused heavily, if not exclusively, on metacognitive EF.

Metacognitive EF refers to the collection of inter-related processes that are involved in maintaining "an appropriate problem-solving set for attainment of a future goal" (Welsh & Pennington, 1988). Although theorists have not yet reached a definitive consensus with respect to which, and how many, particular processes or abilities comprise the metacognitive domain of EF, there has been widespread support for the idea that metacognitive EF includes at least the following abilities: (1) planning ahead and goal setting, (2) initiation, (3) inhibition, (4) cognitive flexibility or shifting, and (5) working memory (Salimpoor & Desrocher, 2006). In neuropsychological assessment, these abilities are often assessed using a handful of performance-based tests that were developed relatively early in the history of the field. For example, the Tower of London

(Shallice, 1982) is used to assess planning; the Controlled Oral Word Association Test (Benton & Hamsher, 1976) to measure initiation; the Stroop Test (Stroop, 1935) to evaluate inhibition; the Trail Making Test, Part B (Army Individual Test Battery, 1944) and Wisconsin Card Sorting Test (Berg & Grant, 1948) to assess cognitive set shifting; and the Digit Span task from the Wechsler scales (Wechsler, 1939, 1949) to test working memory.

The abovementioned performance-based tests have been conceptualized and used rather uncritically as measures of EF for decades. Assuming that the same deficient brain processes that produce poor performance on EF tests will necessarily yield difficulties in situations demanding EF abilities in a person's home, school, or workplace, clinicians have long used clients' scores on these tests to make inferences about their capacities for real-world executive behaviour. More recently, however, this seemingly naïve, one-to-one correlation between performance-based EF test performance and everyday functioning has not only been challenged but actually subjected to empirical investigation. Indeed, studies with various samples of children and adults have shown that scores on performance-based EF tests correlate weakly with behavioural measures of EF and often do not emerge as reliable predictors of real-world functioning. Even when findings are statistically significant, the relationships are only moderate in magnitude (Chaytor, Schmitter-Edgecombe, & Burr, 2006). Not only do such findings cause concern to clinicians who have, for decades, drawn real-world inferences on the basis of EF test performance alone, but they also raise questions about the construct(s) that the performance-based tests assess.

The present study aimed to investigate the degree to which three performance-based measures of EF—namely the Controlled Oral Word Association Test (COWAT; Benton & Hamsher, 1976), the Trail Making Test (TMT; Army Individual Test Battery, 1944), and the Wisconsin Card Sorting Test (WCST; Berg & Grant, 1948)—relate to the Behavior Rating Inventory of Executive Function (BRIEF; Gioia, Isquith, Guy & Kenworthy, 2000), a behavioural measure of EF, in a predominantly Aboriginal/Indian sample of children with Fetal Alcohol Spectrum Disorder (FASD). Considering that executive dysfunction is both common and significant among children with prenatal alcohol exposure, an examination of the relationship between performance-based and behavioural measures of EF is not only interesting but also meaningful in this population.

A thorough appreciation of the present study necessitates an understanding of the following concepts, each of which will be presented below: (1) the nature of Fetal Alcohol Spectrum Disorder (FASD), (2) the conceptualization of EF, its major components, and the measures by which each of these components has traditionally been assessed, (3) the issue of ecological validity with respect to the neuropsychological assessment of EF, and (4) a summary of findings from past studies examining the ecological validity of EF tests in clinical samples. Details about the performance of children with FASD on performance-based and behavioural measures of EF will be integrated where appropriate.

Fetal Alcohol Spectrum Disorder (FASD)

Fetal Alcohol Spectrum Disorder (FASD) is a non-diagnostic umbrella term which denotes the range of outcomes resulting from prenatal alcohol exposure (Riley, Infante, & Warren, 2011). Whereas the most severely-affected individuals on this spectrum are typically diagnosed with Fetal Alcohol Syndrome (FAS) on the basis of (1) facial dysmorphism, (2) growth deficiency, and (3) either structural or functional central nervous system dysfunction, the majority of individuals with histories of prenatal alcohol exposure do not exhibit all of these features and therefore cannot be diagnosed with FAS. Despite their failure to meet diagnostic criteria for FAS, however, these individuals may present with FAS-like physical features, and often have significant cognitive and/or behavioural difficulties that are, in some cases, similar in magnitude to those with FAS (Mattson & Riley, 1998; Mattson, Riley, Gramling, Delis, & Jones, 1998; Rasmussen, 2005; Sampson, Streissguth, Bookstein, & Barr, 2000). In light of this, the Institute of Medicine (Hoyme et al., 2005; Stratton, Howe, & Battaglia, 1996) developed three diagnostic labels to describe those outcomes of prenatal alcohol exposure that are relatively less severe than FAS. Of these labels, Partial Fetal Alcohol Syndrome (pFAS) is diagnosed in cases where there is evidence of some facial characteristics and either growth retardation, CNS deficits, or complex behavioural or cognitive difficulties. Alcohol Related Neurodevelopmental Disorder (ARND) is used for individuals with CNS deficits or complex behavioural or cognitive difficulties. Finally, Alcohol Related Birth Defects (ARBD) describes individuals who present only with some congenital physical abnormalities as a result of prenatal alcohol exposure.

The neurocognitive and behavioural profiles of individuals with FASD are quite variable and, as one might expect, depend on a wide range of factors including the level of prenatal alcohol exposure. Despite this, however, in addition to reduced IQ, learning disabilities, and lower achievement scores, FASD has been associated with primary disabilities in language, attention, working memory, executive function, and socioemotional functioning (Nash, Sheard, Rovet, & Koren, 2008).

While definitive epidemiological data about incidence and prevalence of FASD in Canada are lacking, according to US estimates, about 3 to 6 cases of FASD are found per 1000 live births (Centers for Disease Control and Prevention, 2002). Although considerably higher incidences of FASD, ranging from 25 per 1000 (Asante and Nelms-Matzke, 1985) to 190 per 1000 (Robinson, Conry, & Conry, 1987), have been reported among some ‘high-risk’ First Nations communities (i.e., communities in which FASD was deemed a clear public health concern), this does not appear to be true for all Aboriginal communities: in other tribes, FASD prevalence rates have been found to be comparable to those among non-Aboriginals (Bray & Anderson, 1989).

Research on the alcohol consumption patterns of Aboriginal and non-Aboriginal women is scarce. However, findings in the literature suggest that although Aboriginal women are more likely to abstain from alcohol use as compared to both men and non-Aboriginal women, those who do drink tend to drink heavily (MacMillan et al., 2008; Roberts and Nanson, 2000). For example, in a study comparing the results of the Ontario First Nations Regional Health Survey to those of the National Population Health Survey, First Nations women reported lower rates of alcohol use in the past 12 months (55% vs. 75%). However, of First Nations women who drank, significantly greater proportions reported having five or more drinks on one occasion (43% vs. 24%; MacMillan et al., 2008). The latter pattern of alcohol consumption, sometimes referred to as binge drinking, has been found to be related to the development of FASD (Barr & Streissguth, 2001).

In spite of the commonly-held belief that substance use during pregnancy occurs more frequently among Aboriginal women as compared to non-Aboriginal women (Tait, 2003), a review of the literature on FASD prevalence rates and alcohol consumption patterns among Aboriginals reveals that research in these areas is not only inconclusive but far from complete. At best, the findings published to date highlight the diversity among Aboriginal cultures and indicate a need to gather community-specific information with respect to these issues (Tait, 2003).

Executive Function

Executive function (EF) is an umbrella term for the control or self-regulatory functions that organize and direct all cognitive activity, emotional responses, and overt behaviour (Gioia & Isquith, 2004). From a theoretical standpoint, two types of EF have been differentiated: emotional/motivational EF and metacognitive EF (Fuster, 2001, 2002; Happaney et al., 2004). Emotional/motivational EF, which is largely subserved by the ventromedial prefrontal region of the brain, is responsible for coordinating cognition and emotion/motivation so that an individual can fulfill his basic impulses in a socially appropriate way (i.e., while inhibiting selfish or unsociable behaviours). Executive dysfunction in this domain often affects mood, affect, energy level, initiative, and moral and social behaviour (Anderson, Bechara, Damasio, Tranel, & Damasio, 1999; Barrash, Tranel, & Anderson, 2000; Eslinger & Damasio, 1985; Eslinger, Grattan, Damasio, & Damasio, 1992). Children with impaired emotional/motivational EF, for example, may be apathetic and unmotivated or impulsive and argumentative. In some children, lack of insight and intuition manifest as disregard for social rules and conventions: failing to appreciate the consequences of their actions, they frequently ask embarrassing questions and make socially inappropriate comments (Anderson, 2002).

Metacognitive EF, on the other hand, refers to EF as it is typically understood in neuroscience and neuropsychology: the system that controls and directs the lower-order, domain-specific neuropsychological functions (e.g., language fundamentals, visuospatial functions, and memory) for the purpose of achieving some goal (Gioia & Isquith, 2004). Subserved mainly by the dorsolateral prefrontal cortex, metacognitive EF consists of numerous distinct but inter-related processes including at least the following abilities: (1) planning ahead and goal setting, (2) initiation, (3) inhibition, (4) cognitive flexibility or shifting, and (5) working memory (Donders, 2002; Espy et al., 2002; Gioia, Isquith, Guy, & Kenworthy, 2000; Gioia, Isquith, Kenworthy, & Barton, 2002; Gioia, Isquith, Retzlaff, & Espy, 2002; Griffith, Pennington, Wehner, & Rogers, 1999; Hill, 2004; Ozonoff, 1998; Pennington, Bennetto, McAleer, & Roberts, 1996). Children with executive dysfunction in the metacognitive domain may, for example, have difficulties with planning, organizing, implementing different strategies to a problem, staying on task, or switching between different activities (Anderson, 2002).

In neuropsychology, traditional (i.e., performance-based) assessment of EF almost exclusively involves the evaluation of metacognitive EF. In keeping with this, following this brief introduction is a description of each of five abilities that are widely accepted as metacognitive EF, their implications for everyday behaviour, the performance-based measures that have been employed to assess them, as well as details about how children with FASD perform on these measures.

Planning and goal setting

Planning and goal-setting refer to an individual's ability to set goals (i.e., involving tasks or activities) and to generate the most effective method by which the goals can be attained (Gioia et al., 2000; Salimpoor & Desrocher, 2006). This involves conceptualizing an end state; generating alternative courses of action to reach the end state; choosing one of the courses of action; identifying the steps, skills and materials relevant to the chosen course of action; correctly sequencing these identified steps; and finally, anticipating any problems that may come up in the process of goal attainment (Gioia et al., 2000; Lezak, Howieson, Loring, Hannay, & Discher, 2004; Ozonoff, 1998; Salimpoor & Desrocher, 2006). In everyday life, efficient planning and goal-setting allows for the timely accomplishment of daily tasks and the generation of alternative sequences of action when an initial routine is unexpectedly interrupted (Salimpoor & Desrocher, 2006).

Planning and goal-setting abilities are typically evaluated using tasks that present the patient with a predetermined goal and require him or her to accomplish the goal by developing and implementing a strategy that is as accurate and efficient as possible (Salimpoor & Desrocher, 2006). Tower tasks, in which patients are asked to move a number of rings from a starting peg to a goal peg in as few moves as possible and under specific constraints (e.g., a larger ring cannot be placed on top of a smaller one), are often used for this purpose. Examples of such tests include the Tower of London (Shallice, 1982), Tower of Hanoi (Boyrs, Spitz, & Dorans, 1982; Welsh & Huizinga, 2001), Progressive Planning Test (Kodituwakku, 1993), Stockings of Cambridge subtest from the Cambridge Neuropsychological Test Automated Battery (Robbins et al., 1994), and the Tower Test from the Delis-Kaplan Executive Function System (Delis, Kaplan, & Kramer, 2001). In addition to variants of the tower task, maze completion tasks such as the Porteus Mazes (Krikorian & Bartok, 1998; Porteus, 1959) have also been used to assess planning ability. In these tasks, individuals are asked to draw paths through mazes of varying complexity while obeying specific rules.

Alcohol-exposed children have been observed to perform poorly on a number of different tower tasks. For example, on the Progressive Planning Test, they solved fewer problems than control children, obtained lower total scores, and perseverated on incorrect strategies (Aragon et al., 2008; Kodituwakku, Handmaker, Cutler, Weathersby, & Handmaker, 1995b; Mattson et al., 2010). Similarly, on the Stockings of Cambridge subtest of the CANTAB, children with FASD spent less time than controls when planning their strategies to solve the problems and, not surprisingly, also solved fewer problems in the minimum number of moves allowed (Green et al., 2009). Finally, on the Tower Test of the D-KEFS, alcohol-exposed children used more moves to solve the problems, violated rules more frequently, and passed fewer items overall when compared to the normative mean and control children (Astley et al., 2009; Mattson, Goodman, Caine, Delis, & Riley, 1999). They also violated the rules more often than controls on this task (Astley et al., 2009). The poor performance of children with FASD across different tower tasks suggests that these children have significant difficulty with planning.

Initiation

Initiation refers to one's ability to (1) develop a mental set or course of action through the independent generation of ideas, alternatives, and problem-solving strategies, and (2) execute the mental set by beginning some task or activity (Gioia et al., 2000; Salimpoor & Desrocher, 2006; Turner, 1997). In everyday life, effective initiation allows an individual to imagine and implement abstract or multiple solutions to a problem (Salimpoor & Desrocher, 2006).

Initiation has typically been assessed by tests of fluency in which individuals are asked to generate as many different verbal (e.g., words beginning with a specific letter of the alphabet or words belonging to a single semantic category) or nonverbal (e.g., novel designs) responses as possible within a time limit (Salimpoor & Desrocher, 2006). Examples of such tasks include the Controlled Oral Word Association Test (Benton & Hamsher, 1976), the Ruff Figural Fluency Test (Vik & Ruff, 1988), as well as the Verbal Fluency and Design Fluency tests from the D-KEFS (Delis et al., 2001).

Initiation abilities of children with FASD have been examined using both verbal and nonverbal fluency tasks. Verbal fluency tests, such as the COWAT and the Verbal Fluency Test from the D-KEFS, take the form of phonemic (i.e., letter) fluency tasks and semantic (i.e., category) fluency tasks. While alcohol-exposed children have shown deficits in both letter and category fluency performance in some studies (Mattson et al., 2010; Schonfeld, Mattson, Lang, Delis, & Riley, 2001; Vaurio, Riley, & Mattson, 2008), in other studies, their deficits have been limited to letter fluency (Aragon et al., 2008; Kodituwakku et al., 1995b; Rasmussen & Bisanz, 2009). With respect to nonverbal fluency, two studies have examined the performance of alcohol-exposed children using the Design Fluency subtest from the D-KEFS with mixed results. While alcohol-exposed children obtained lower scores on this subtest than the normative mean in one study (Schonfeld et al., 2001), in the other, their performance was at par with a control group (Rasmussen & Bisanz, 2009). The discrepancy in these results appears to be due to the use of different comparison samples in the two studies (Mattson, Crocker, & Nguyen, 2011). Taken together, results from fluency measures indicate that children with FASD have difficulty with initiation at least in the verbal domain.

Inhibition

Inhibition refers to the ability to suppress irrelevant distractions in favour of maintaining an already-activated mental set (Salimpoor & Desrocher, 2006). Behaviourally, this may entail resisting an impulse or a prepotent response, or ceasing a behaviour at the appropriate time (Bialystok & Martin, 2004; Gioia et al., 2000; Levin, Hanten, Zhang, Swank, & Hunter, 2004).

The Stroop Test (Stroop, 1935) and its variants, such as the Color-Word Interference test from the D-KEFS (Delis et al., 2001), have been employed to assess patients' capacities for inhibition. In this task, the patient views colour names printed in coloured ink, and is asked to name the colour of the ink. On some trials, the printed colour name and the colour of the ink are congruous (e.g., the word 'green' printed in green ink), while on other trials, they are incongruous (e.g., the word 'green' printed in red ink). Thus, successful performance on the incongruous trials requires inhibition of the prepotent response, which is reading the word. The patient's score on this test, derived by subtracting his or her reaction times on the congruent conditions from the incongruent conditions, is thought to reflect his or her capacity for inhibition (Salimpoor & Desrocher, 2006).

On variants of the Stroop Test, alcohol-exposed children have been found to make more errors when compared to both the normative mean and to a group of healthy controls (Mattson et al., 1999; Rasmussen & Bisanz, 2009), indicating deficits in inhibition.

Cognitive flexibility

Cognitive flexibility (i.e., shifting) is the ability to shift flexibly from one mental set to another in accordance with the demands of the environment (Salimpoor & Desrocher, 2006). More specifically, it involves monitoring environmental cues and utilizing this feedback to make appropriate mental, attentional, and behavioural transitions (Anderson, Damasio, Dallas, & Tranel, 1991; Rothke, 1986; van der Sluis, de Jong, & van der Leij, 2004). In daily life, effective shifting underlies an individual's ability to move freely from one situation, activity, or aspect of a problem to another as requirements change (Gioia, Isquith, Guy, & Kenworthy, 2000).

During neuropsychological testing, shifting has most commonly been assessed with tasks requiring cards or objects to be sorted on the basis of various rules (Salimpoor & Desrocher, 2006). The most popular of these tasks is the Wisconsin Card Sorting Test (Berg & Grant, 1948). In this task, cards can be matched based on colour, form, or number, and the criteria for card matching changes, without warning to the patient, after a certain number of correct responses. Thus, when a previously-reinforced cognitive set begins to result in an incorrect answer, the patient must exercise cognitive flexibility in order to generate a new mental set. Persistence on the "old" card-sorting rule yields perseverative errors, which are taken as evidence of cognitive inflexibility. Another test that has been used to assess set-shifting abilities is Part B of the Trail Making Test (Battery, 1944; Reitan, 1958), which requires patients to connect 25 encircled numbers and letters in alternating order as quickly as possible. A variant of this test has also been included in the D-KEFS (Delis et al., 2001).

On card sorting tasks, children with FASD have been observed to make more errors, both perseverative and otherwise, as compared with healthy controls (Astley et al., 2009; Kodituwakku, Handmaker, Cutler, Weathersby, & Handmaker, 1995a; Kodituwakku, May, Clericuzio, & Weers, 2001; McGee, Schonfeld, Roebuck-Spencer, Riley, & Mattson, 2008; Olson, Feldman, Streissguth, Sampson, & Bookstein, 1998; Vaurio et al., 2008). Their higher numbers of non-rule-based errors (Olson et al., 1998) and lower percentages of conceptual-level responses (McGee, Schonfeld, et al., 2008; Vaurio et al., 2008) suggest that they tend to respond in a disorganized and unplanned manner. Consistent with these findings, children with FASD have been found to complete fewer categories on the WCST than their neurotypically-developing peers (Coles et al., 1997; Kodituwakku et al., 1995b; Kodituwakku et al., 2001; Olson et al., 1998).

On Part B of the Trail Making Test, alcohol-exposed children performed significantly worse than healthy controls (Vaurio et al., 2008). Similarly, on the number-letter switching condition of the Trail Making Test from the D-KEFS, they not only demonstrated longer times to completion but also made more errors than control children (Astley et al., 2009; Mattson et al., 1999; Mattson et al., 2010; Rasmussen & Bisanz, 2009).

Working memory

Working memory is the ability to hold new information in mind long enough to allow for manipulation, problem solving, or the accomplishment of some task (Gioia, Isquith, Guy, et al., 2000; Kimberg, 1996; Pennington, 1994; Salimpoor & Desrocher, 2006). In daily life, working memory is essential to remembering the rules of an ongoing activity or task, keeping track of one's own prior behaviours, manipulating information mentally, executing multi-step activities, and following complex instructions (Gioia et al., 2000; Salimpoor & Desrocher, 2006).

Verbal working memory is typically assessed using the Digit Span task from the Wechsler Intelligence scales (Wechsler, 1939, 1949, 1991, 2004), which requires the repetition of strings of digits in forward and backward order. Nonverbal working memory, on the other hand, is evaluated using the Self-Ordered Pointing Task (Petrides & Milner, 1982) and the Spatial Working Memory subtest of the CANTAB (Robbins et al., 1994). In these tasks, patients are to point to one picture per page in a series of pages, while ensuring that they do not point to the same picture more than once. The Spatial Span subtest of the CANTAB, in which patients must match a sequence they just observed by pointing to a series of squares on a screen, has also been used to assess nonverbal working memory.

Alcohol-exposed children have been found to recall fewer digits than do controls on the backwards condition of the Digit Span task (Aragon et al., 2008; O'Hare et al., 2009; Olson et al., 1998). With respect to spatial working memory, children with FASD made more errors than controls on the Spatial Working Memory subtest of the CANTAB

(Green et al., 2009; Mattson et al., 2010). These children also demonstrate lower spatial span lengths than their peers (Mattson et al., 2010).

Taken together, the studies reviewed above indicate that, when evaluated using traditional performance-based tests of EF, children with FASD exhibit impairments in all major subdomains of metacognitive EF (i.e., planning, initiation, inhibition, cognitive flexibility, and working memory).

The Issue of Ecological Validity in EF Assessment

Ecological validity can be defined as the extent to which test results are generalizable to "naturally occurring" events in the real world (Brunswik, 1955). According to Chaytor and Schmitter-Edgecombe (2003), two concepts, namely verisimilitude and veridicality, underlie the ecological validity of any neuropsychological test. Of these, verisimilitude reflects the degree of similarity between the nature of the test and testing conditions, on the one hand, and the demands of the patient's daily life, on the other. It asks the question, 'to what extent does a task in which a patient is required to rapidly transcribe symbols from a legend resemble one or more activities that he or she does at home, at school, or at work?' The second concept, veridicality, refers to the extent to which performance on a test actually predicts some aspect of the patient's daily functioning. When either theory or reason suggests that a test might predict real-world behaviour, the relationship between scores on performance-based tests and measures of everyday functioning (e.g., behaviour rating scales or inventories) can be examined empirically.

Verisimilitude of EF tests

Neuropsychological tests of EF do not fare well in terms of verisimilitude: It is not often, at home, school, or work, that we are asked to sort cards based on unexpectedly-changing rules or to identify the colour of text that actually spells a different colour name. In fact, verbal fluency tasks, which require the generation of words that either begin with the same letter of the alphabet or belong to the same semantic category, may constitute a single exception to this trend. Interestingly, however, the low verisimilitude of EF tests is largely intentional. Shallice (1990) has argued that because individuals can often perform simple and routinized tasks rather instinctively, such tasks do not activate an individual's EF abilities. In order to ensure that EF tests are, in fact, drawing upon EF processes, EF-test developers must maximize creativity and novelty, and create tasks that involve the formulation, implementation, and evaluation of new plans and strategies. Considering the fact that a novel task will, by nature, have little, if any, resemblance to the demands of an individual's daily life, it appears that verisimilitude must be sacrificed for the assessment of EF.

Another criticism with respect to the verisimilitude of EF tests is the fact that most of these tests require overly simplistic responses by the patient. In neuropsychological testing, evidence of executive function, or conversely executive dysfunction, is taken from tasks in which a patient physically sorts cards from a deck into four piles; presses keys on a computer keyboard; repeats or manipulates certain words or digits; or draws a line to connect circles that are scattered on a page. In the real world, however, executive function or dysfunction manifests as success or failure in planning, initiating, and completing various daily activities (e.g., a child's accomplishing his morning routine or completing his homework) while inhibiting distractions yet remaining flexible in the event that priorities or environmental demands change. Clearly, there is a significant discrepancy in the complexity of those behaviours that are considered indicative of executive function or dysfunction in the testing situation and those that reflect these processes in real life. This discrepancy contributes further to the low verisimilitude of EF tests.

In addition to test demands, the highly-structured and interactive conditions of neuropsychological testing represent another barrier to the verisimilitude of EF tests (Sbordone, 2000). In fact, it has been argued that such conditions may actually be detrimental when it comes to the assessment of EF in particular. Whereas neuropsychological tests are intended to be administered in a quiet room with minimal distraction, in everyday life, an individual’s neuropsychological abilities are expressed in complex, noisy, and unstructured environments (Sbordone, 2010). Considering that executive function (and dysfunction) is context-dependent, the failure of the testing conditions to simulate the complexity of the real world likely renders the detection of a patient’s executive difficulties more difficult (Sbordone & Guilmette, 1999). Furthermore, in a neuropsychological evaluation, the patient works one-on-one with a pleasant and supportive examiner who is not only structuring and initiating activities for him but also encouraging him to put forth his full effort throughout the assessment. Because such one-to-one support is uncommonly the case in a patient’s real-life settings, his or her difficulties may not become evident during neuropsychological testing. In fact, it has been argued that through such an active role in testing, the examiner may actually be serving as an external executive control system for the patient (Stuss & Alexander, 2000; Stuss & Benson, 1986), further complicating the detection of executive dysfunction during the assessment.

Improving verisimilitude in EF assessment. Acknowledging the low verisimilitude of performance-based EF tests, neuropsychologists have turned to a relatively new method of EF assessment: behaviour rating scales. High in verisimilitude, these scales typically contain a wide range of items that reflect various behavioural manifestations of EF in a patient's home, school, or work environment. The rater, who could be a caregiver, teacher, or the patient himself, is asked to endorse the frequency with which these behaviours occur.

Although several behaviour inventories have been developed over the years, the Behavior Rating Inventory of Executive Function (BRIEF; Gioia, Isquith, & Kenworthy, 2000) is one of the most popular measures for ecologically-valid assessment of EF in children. The 86-item inventory comes in both parent- and teacher-rating forms, and provides scores on eight clinical scales, each reflecting a different component of EF: Inhibit, Shift, Emotional Control, Initiate, Working Memory, Plan/Organize, Organization of Materials, and Monitor. Scores from these scales combine to form two broader index scores, the Behavioural Regulation Index (BRI) and Metacognition Index (MI), and an overall score called the Global Executive Composite (GEC).

Behaviour ratings of EF among children with FASD. Several studies have examined BRIEF profiles of children with FASD, as rated by their caregivers. Results from these studies show that alcohol-exposed children consistently obtain scores not only reflecting poorer executive functioning than controls on all scales and summary indices of the BRIEF, but that their scores are often in the clinically-significant range (Astley et al., 2009; McGee, Fryer, Bjorkquist, Mattson, & Riley, 2008; Rasmussen, McAuley, & Andrew, 2007). Despite their poor performance across BRIEF scales, however, it appears that children with FASD struggle with certain aspects of EF more than others. They tend to obtain highest scores, indicating more difficulty, on the Working Memory scale (Rasmussen, Horne, & Witol, 2006; Rasmussen et al., 2007), and lowest scores, indicating least difficulty, on the Organization of Materials scale (McGee, Fryer, et al., 2008; Rasmussen et al., 2006; Rasmussen et al., 2007). Interestingly, alcohol-exposed children with and without FAS, the latter representing the most severe form of the disorder, obtain similar scores on the BRIEF, suggesting comparable impairments in EF among these groups despite severity (Astley et al., 2009; Chasnoff, Wells, Telford, Schmidt, & Messer, 2010; Rasmussen et al., 2006).

Veridicality of EF tests

Veridicality, the second component of ecological validity, is defined as the ability of a test to predict some aspect of a patient's everyday functioning. As discussed above, the EF system directs the lower-order domain-specific neuropsychological functions for the purpose of achieving some external goal. Conceptualizing life activities as a series of external goals, both big and small, it has been argued that it may be more meaningful to investigate the veridicality of EF tests than the veridicality of tests of other neuropsychological domains (Mitchell & Miller, 2008).

Ecological validity studies examine the relationship between neuropsychological test scores and measures of everyday functioning. With respect to the assessment of EF, therefore, such studies can be used to investigate the relationship between the traditional or classical EF tests (e.g., Wisconsin Card Sorting Test, Stroop Test, Tower of London) and behavioural measures of EF. Although various behavioural measures of EF can be used for this purpose (e.g., observation of the patient during simulated everyday tasks, clinical rating scales, as well as self-report and informant questionnaires), these methods are not without their limits (Chaytor et al., 2006). For example, comparing EF test performance to observations from simulated tasks may yield higher ecological validity due, in part or in whole, to elements that are common to both settings (e.g., artificial situation, stress of being evaluated). Patient self-reports may also be unreliable, especially in cases of executive dysfunction, where lack of insight and deficient self-monitoring are likely. Considering these issues, researchers have largely relied on informant rating scales as measures of everyday EF in ecological validity studies of EF tests.

The above discussion, which outlines the approach by which the ecological validity of neuropsychological tests is typically examined, sets the stage for a review of the existing literature on the ecological validity of the traditional, performance-based EF tests. Although veridicality specifically refers to the *prediction* of everyday behaviour from neuropsychological test scores (i.e., using multiple regression analysis), studies utilizing correlational analyses to examine the relationship between EF tests and behavioural measures of EF constitute a large majority of the literature in this area and, thus, are included in the review. Furthermore, the review of the literature is limited to studies (1) which were conducted with children or adolescents, and (2) in which EF tests were compared to informant-based behavioural ratings of EF processes in particular. The following clinical samples are considered: traumatic brain injury (TBI), other neurological conditions, attention-deficit hyperactivity disorder (ADHD), and autism spectrum disorders (ASD).

Brain injury. Vriezen and Pigott (2002) investigated the relationship between performance-based tests of EF and parent ratings on the Behavior Rating Inventory of Executive Function (BRIEF) in a sample of 48 children with moderate-to-severe TBI. Neuropsychological measures included the WCST, TMT-B, letter fluency (F, A, S), and category fluency (animals). Interestingly, none of the correlations between performance-based test scores and BRIEF index and composite scores reached significance. In another study, Conklin, Salorio, and Slomine (2008) examined the ecological validity of the backwards condition of the Digit Span task from the Wechsler intelligence scales, which has long been used as a measure of working memory, in a sample of 62 children with moderate-to-severe TBI. They found the children's scores on this task were not significantly correlated with their scores on the Working Memory scale of the BRIEF. Furthermore, performance on the backward condition of Digit Span was not a significant predictor of scores on the Working Memory scale of the BRIEF specifically. In contrast to these studies, in which data from performance-based tests and behavior ratings was obtained at the same point in time, Maillard-Wermelinger et al. (2009) investigated whether scores on performance-based EF tests, which were administered shortly after the children acquired mild TBIs, predicted later ratings on the BRIEF. In this study, the Spatial Working Memory and Stockings of Cambridge subtests from the Cambridge Neuropsychological Test Battery (CANTAB), which are essentially computerized versions of the self-ordered pointing task and tower task respectively, were used to assess EF within three weeks of injury. BRIEF ratings were obtained at the initial assessment (retrospectively, to reflect premorbid functioning), and at three- and 12-months post-injury. Hierarchical regression analyses revealed that, after controlling for group membership, age at injury, and initial (premorbid) ratings on the BRIEF, the two CANTAB subtests, together and individually, accounted for significant variance in several BRIEF scales at three months post-injury and in most scales at 12 months post-injury, although the relationships were modest in magnitude. Notably, the Spatial Working Memory subtest was a significant predictor of scores on the Working Memory scale of the BRIEF at 12 months post-injury. Taken together, these findings suggest that, for children with histories of TBI, scores on performance-based EF tests are generally not related to BRIEF scale or index scores as concurrently rated by caregivers. However, their performance on EF tests administered shortly after TBI may be predictive of the "everyday" executive difficulties they will experience later on.

Neurological conditions. Anderson, Anderson, Northam, Jacobs, and Mikiewicz (2002) examined the ecological validity of several performance-based tests of EF in a mixed sample of 44 children with early-treated phenylketonuria, 45 with early-treated hydrocephalus, 20 with frontal focal lesions, and 80 control children. Traditional performance-based tests of EF included the Tower of London (TOL), and the Controlled Oral Word Association Test (COWAT), while the BRIEF was used to assess everyday executive functioning. Although scores on the TOL were not significantly correlated with any BRIEF parameters, weak correlations were found between the participants' total scores on the COWAT and the BRIEF Working Memory (.30), Inhibit (.29), and Emotional Control (.24) scales. In another study, MacAllister et al. (2012) examined the relationship between scores on the TOL and the BRIEF in a sample of 87 children with pediatric epilepsy. They found that compared to the normative sample, children with epilepsy obtained lower total move scores, total correct scores, made more rule-violation errors, and had longer total problem solving times. Their T-scores on all BRIEF scales and indexes were also higher than those of the normative sample. However, despite the fact that both of these measures provided evidence of executive dysfunction in this sample, none of the correlations between TOL scores and BRIEF parameters reached significance. Finally, Parrish et al. (2007) investigated the ecological validity of two tasks from the D-KEFS, which closely resemble traditional performance-based EF tests, in 53 children with epilepsy. The Sorting Test, previously called the California Card Sorting Test, is a card-sorting task that emphasizes cognitive flexibility, whereas the Color-Word Interference Test, which is similar in nature to the Stroop task, assesses inhibition. Both the Sorting Test and Color-Word Interference Test were found to be modestly correlated (-0.28 and -0.33, respectively) with the BRIEF Metacognition Index, with lower scores on the performance-based tasks associated with higher (i.e., worse) scores on the BRIEF. Furthermore, the BRIEF Metacognition Index was a significant predictor of scores on both of the D-KEFS subtests. Taken together, these studies provide mixed evidence for the ecological validity of traditional EF tests in children with various neurological conditions. Whereas verbal fluency (initiation), card-sorting (cognitive flexibility), and color-word interference (inhibition) tasks have been found to be weakly related to behavioural measures of EF, the TOL has not shown a significant association, which is particularly concerning as tower tasks are routinely used to assess planning ability.

Attention-Deficit Hyperactivity Disorder. Toplak, Bucciarelli, Jain, and Tannock (2009) investigated the ecological validity of various traditional EF tests in a sample of 46 adolescents with ADHD and 44 comparison controls. Traditional neuropsychological measures of EF included Part B of the Trail Making Test, the Digit Span task from the

WISC-III (Wechsler, 1991), the Spatial Span task from the WISC-III Process Instrument (Kaplan, Fein, Kramer, Delis, & Morris, 1999), and the Stockings of Cambridge subtest from the CANTAB, which is essentially a computerized version of the Tower of London. The Inhibit, Shift, Working Memory, and Plan/Organize indexes of the BRIEF (as rated by parents and teachers) were used as measures of everyday EF behaviour. Many significant correlations, weak-to-moderate in magnitude, were found between performance-based measures and parent- and teacher-ratings on the BRIEF. Specifically, Part B of the Trail Making Test was significantly correlated with the Inhibit (.29), Working Memory (.39), and Plan/Organize (.34) scales on the parent report form, and the Working Memory (.31) scale on the teacher form. A Working Memory composite score, derived from the Digit Span and Spatial Span tasks, was significantly correlated with the Inhibit (-.30), Shift (-.40), Working Memory (-.41), and Plan/Organize (-.37) scales of the parent report form, and the Working Memory (-.33) scale of the teacher report form. Finally, scores on the Stockings of Cambridge subtest of the CANTAB were significantly correlated with the Inhibit scale (-.34) of the teacher report form of the BRIEF. While these results provide some evidence of ecological validity of performance-based EF tests, it is notable that these tests were not uniquely associated with their respective scales on the BRIEF. For example, Part B of the Trail Making Test, which is thought to be a measure of shifting, was not uniquely associated with the BRIEF Shift scale.

Autism Spectrum Disorder. Teunisse et al. (2012) compared performance on a shortened version of the Wisconsin Card Sorting Test (WCST-S) with two parent-rated measures of flexibility, the Shift scale of the BRIEF and the Behavior Flexibility Rating Scale-Revised (BFRS-R), in a sample of 20 children with Autism Spectrum Disorders

(ASD). Although scores on the two rating scales were significantly correlated with each other, the WCST-S was not significantly correlated with either of them.

Several studies have investigated the veridicality of the traditional, performance-based measures of EF in various clinical samples of children and adolescents, including TBI, neurological diseases, ADHD, and ASD. Results from these studies provide little support for a relationship between performance-based and behavioural measures of EF. Indeed, both correlational and regression analyses often fail to yield significant results and, when they do, the observed relationships are only moderate in magnitude. Considering that neuropsychologists have made inferences about patients' real-world functioning on the basis of EF test performance for decades, such findings are not only disappointing but also concerning to the clinician. Having said this, however, it is important to note that past research in this area is not without its limitations. First, some studies have involved small sample sizes, rendering the detection of an effect that is likely small in magnitude to begin with, even more difficult. Second, researchers have employed limited statistical procedures to investigate the ecological validity of EF tests. For example, while Pearson correlations between EF test performance and scores on EF rating scales have been examined in many studies, few have taken advantage of multiple regression analysis. Multiple regression analysis allows researchers to determine how well a *combination* of two or more variables predicts the criterion variable. Because neuropsychologists often administer more than one EF test during an assessment, it would be meaningful to investigate the combined ability of several EF tests to predict scores on a behavioural measure of EF. Finally, a weak relationship between EF tests and EF inventories has been attributed, by several researchers, to the idea that these measures may not be evaluating the same construct(s). Despite the availability of various statistical methods to investigate the latter hypothesis (e.g., factor analysis, canonical correlation analysis), few studies, only one of which involved a pediatric sample, have examined the nature of the relationship between performance-based and behavioural measures of EF. In their study, Bakar, Taner, Soysal, Karakas, and Turgay (2011) conducted a factor analysis of the BRIEF in a sample of boys with ADHD, and found that the Stroop Test, Wisconsin Card Sorting Test, and Raven Standard Progressive Matrices did not load on the same factors as the BRIEF. While these results provide preliminary evidence that EF tests may, in fact, be measuring different constructs in children than those they have traditionally been thought to measure, more studies are required to confirm this. Considering these limitations, it is clear that more rigorous investigation of the ecological validity of EF tests is not only possible but necessary.

Present Study

The purpose of the present study was to advance the literature on the veridicality of traditional, performance-based tests of EF. While studies have shown that alcohol-exposed children (1) demonstrate impairments on several neuropsychological tests of EF, and (2) frequently obtain elevated scores across all scales and indexes of the BRIEF when rated by caregivers, the relationship between performance-based EF measures and the BRIEF has not been empirically investigated in this population. The present study aimed to accomplish this.

Of primary interest in this study was the extent to which performance-based EF tests (WCST, COWAT, and TMT) relate to the BRIEF. This was evaluated in two ways. First, the scores of alcohol-exposed children on EF tests were correlated with their scores on the eight scales and two indexes of the BRIEF. Based on the literature, it was expected that these correlations, although moderate in magnitude, would be statistically significant. Second, two multiple regressions were used to determine whether scores from performance-based tests of EF could, in combination, predict scores on the Metacognition and Behavior Regulation Indexes of the BRIEF, respectively. The combination of scores from the three performance-based tests was expected to result in a statistically significant regression model, with at least one performance-based EF test emerging as a significant predictor in each of the two analyses.

In addition to examining the extent of the relationship between the neuropsychological and behavioural measures of EF in a sample of children with FASD, the present study investigated the nature of the relationship using a canonical correlation analysis (CCA). CCA is a statistical technique which allows for the association between the neuropsychological measures of EF and the eight scales of the BRIEF to be identified and measured (Stevens, 2009). Because the latter analysis served as an exploratory procedure, it was not accompanied by an *a priori* hypothesis.

CHAPTER 2

METHOD

Participants

This study was conducted using archival data from a clinical database containing approximately 800 consecutive cases of children, generally referred by child psychiatrists, pediatricians, pediatric neurologists, and social workers (social services), to a child neuropsychologist in private practice in Edmonton, Alberta, Canada. At the time of each assessment, consent was obtained for the child's demographics, diagnostic information, and neuropsychological test data to be used for future archival research. Information unique to each participant was coded and all identifying information was removed before the database was released to the researcher for this study.

Of the cases in the database, children were selected for inclusion in the present study on the basis of (1) diagnosis (i.e., they had to have been assigned a primary diagnostic grouping of "FASD" by the licensed clinical neuropsychologist conducting the assessment), (2) age between nine and 16 years, and (3) the availability of scores on the BRIEF.

Many of the alcohol-exposed children in the group used for this study had been diagnosed with some variant of FASD (e.g., FAS, pFAS, ARND) prior to their neuropsychological assessment; the purpose of the assessment in these cases was to determine the children's current levels of functioning and to establish clinical pictures of strengths and weaknesses in order to assist with future planning of services, supports, and programs that would be most effective in helping them to succeed. In the remaining cases, the clinician was asked to establish or confirm a diagnosis on the fetal alcohol spectrum. The Seattle Model, which uses information about growth factors, facial morphology, central nervous system dysfunction, and an admission of alcohol use during pregnancy to arrive at diagnoses on the fetal alcohol spectrum, was used for this purpose.

The final sample used in the present study consisted of 96 children with FASD, 62.5% of whom were male. All of the children were assessed between January 2003 and January 2014. While the majority of children were Aboriginal/Indian (78.1%), the sample also included children identified as Caucasian (16.7%), Black (1.0%), and Other (2.1%). Information about race was unavailable for two participants in the sample. Further information pertaining to the sample (i.e., age, education, and estimates of intellectual functioning) is provided in Table 1.

Table 1

Mean age, education, and intellectual functioning estimates of children in sample

Estimate	*n*	*M*(*SD*)	Range
Demographic variables			
Age	96	13.30(2.37)	9.0 – 16.0
Education	95	7.86(2.49)	2.0 – 12.0
Intellectual functioning[a]			
Verbal IQ	94	76.69(12.56)	47.0 – 108.0
Performance IQ	94	86.27(14.39)	52.0 – 122.0
Full Scale IQ	94	76.76(12.65)	41.0 – 102.0

[a]Estimates of intellectual functioning are based on the WISC-III and, following its release, the WISC-IV. Scores from the WISC-IV Verbal Comprehension Index and Perceptual Reasoning Index were substituted for the WISC-III Verbal IQ and Performance IQ, respectively.

In light of the fact that Aboriginal children constituted a large majority of the sample used in the present study, further information was obtained from the source of the database with respect to these children's living arrangements and educational placements (L. Flaro, personal communication, September 2, 2014). Although precise statistics were not available, the Aboriginal children in the FASD sample were usually raised on reserves until they were apprehended. Following apprehension, most of the children were placed equally with Caucasian and Aboriginal families, in homes off of the reserve: whereas approximately 20 to 30% of the alcohol-exposed children were adopted, the remaining 70 to 80% were placed in foster homes. In the latter group, the children were usually either moved between foster homes or placed in group homes, depending on their behavioural and emotional functioning. With respect to language proficiency and educational placements, the Aboriginal children in the sample were usually exposed to the English language from birth, and most of them attended either regular or specialized school programs.

While the consequences of prenatal alcohol exposure constituted the primary disability for all of the children included in the study (as indicated by a primary diagnostic grouping of "FASD"), comorbid disabilities were identified for a considerable number of children in the sample. Relevant secondary diagnostic groupings, along with the percent frequency of each, are reported in Table 2.

Table 2

Frequencies of secondary diagnoses among children in sample

Secondary diagnostic grouping	Percent
Learning Disability	13.5
Asperger Syndrome	10.4
Nonverbal Learning Disability	7.3
Personality Disorder	4.2
Attention-Deficit/Hyperactivity Disorder	3.1
Bipolar Disorder	3.1
Posttraumatic Stress Disorder	2.1
Anxiety Disorder	1.0
Executive dysfunction	1.0
Fledgling Psychopath	1.0
Intellectual Disability	1.0
Language Impairment	1.0
Mood disorder	1.0
Pica	1.0
Reactive Attachment Disorder	1.0
Schizophrenia	1.0
Traumatic Brain Injury	1.0

Note. Only 55.2% of the children in the sample were assigned secondary diagnostic groupings. For the remaining 44.8% of the sample, only a primary diagnostic grouping of FASD was indicated.

Measures

Although the children in this sample typically underwent comprehensive neuropsychological evaluations, only the measures relevant to the present study are described below.

Wechsler Intelligence Scale for Children—Third and Fourth Editions (Wechsler, 1991; 2004). General intellectual functioning was assessed using the Wechsler Intelligence Scale for Children—Third Edition (WISC-III; Wechsler, 1991) and, shortly after its release, the Wechsler Intelligence Scale for Children—Fourth Edition (WISC-IV; Wechsler, 2004). The WISC-III and WISC-IV each contain 10 core subtests that assess various reasoning abilities, working memory, and processing speed.

While both measures provide a Full Scale IQ (FSIQ; an overall measure of intellectual function), they differ slightly in terms of their index scores. The WISC-III provides index scores for Verbal IQ (VIQ) and Performance IQ (PIQ), whereas the WISC-IV provides scores for the Verbal Comprehension Index (VCI), Perceptual Reasoning Index (PRI), Working Memory Index (WMI), and Processing Speed Index (PSI).

For the purposes of the present study, the FSIQ, VIQ, and PIQ were utilized as estimates of intellectual functioning. In cases where the WISC-IV was administered, scores on the VCI and PRI were substituted for VIQ and PIQ respectively, in accordance with guidelines provided with the WISC-IV Administration and Technical manuals (Wechsler, 2004). It is noteworthy that these indexes differ in composition and therefore are not perfectly interchangeable. The correlation between FSIQ scores from the two measures is .89; the correlation between VIQ and VCI is .83; and the correlation between PIQ and PRI is .73 (Sattler, 2008).

Behavior Rating Inventory of Executive Function (BRIEF; Gioia, Isquith, & Kenworthy, 2000). As described above, the BRIEF is a rating scale that assesses a wide range of everyday behaviours which differentially emphasize various executive processes. These executive processes are reflected in the measure's eight theoretically- and empirically-derived scales: Inhibit, Shift, Emotional Control, Initiate, Working Memory, Plan/Organize, Organization of Materials, and Monitor. Of these, the former three scales comprise the Behavior Regulation Index (BRI), while the latter five scales form the Metacognition Index (MI). The BRIEF also yields an overall score called the General Executive Composite (GEC). All of the BRIEF scores are age- and gender-standardized with a mean of 50 and a standard deviation of 10, so that higher scores represent greater executive dysfunction.

The parent form of the BRIEF was administered to the caregivers of all of the participants in the present study. Internal consistency for this form ranges from.80 to .98 (Gioia, Isquith, Retzlaff, & Pratt, 2001). *T* scores on the eight scales (Inhibit, Shift, Emotional Control, Initiate, Working Memory, Plan/Organize, Organization of Materials, and Monitor) and two indexes (MI and BRI) were used in the analyses.

Wisconsin Card Sorting Test (WCST; Heaton, Chelune, Talley, Kay, & Curtiss, 1993). On the WCST, a child is asked to match key cards from a deck, one by one, to one of four stimulus cards. The cards can be matched based on one or more of three criteria (colour, form, and number), and the child must use feedback from the examiner to determine which of these matching criteria is correct at any given time. Once the child demonstrates an understanding of the operating matching criterion by obtaining ten consecutive correct scores, the criterion is changed without warning to the child. Thus, the child must monitor his or her performance and shift cognitive set in order to continue to be successful. The test is discontinued when the child successfully completes six categories (10 consecutive correct matches per category) or when all 128 cards are sorted.

On the WCST, set-shifting ability is primarily captured by a child's perseverative errors. A perseverative error represents a specific type of error which results when a child continues to match cards by the previously-correct criterion in spite of feedback indicating that his or her responses are incorrect. This score has been found to be fairly reliable (generalizability coefficient = .52, SE_M = 10.39).

It is noteworthy that the participants in the present study were administered a computerized version of the WCST. According to Tien, Spevack, Jones, Pearlson, Schlaepfer, and Strauss (1996), however, performance on perseverative errors does not differ between the computerized and manual versions of the WCST. Considering this, the perseverative error raw scores of children in the present study were standardized to *T* scores using the age-stratified normative data provided in the Heaton et al. (1993) test manual, such that higher *T* scores represented better performance or fewer perseverative errors.

Controlled Oral Word Association Test (COWAT; Benton & Hamsher, 1976). The COWAT is a task of phonemic verbal fluency in which children are asked to generate orally as many words as they can that begin with the letters F, A, and S, respectively, in three 60-second trials. Children are not permitted to use proper nouns (i.e., names of people or places). Test-retest reliability for the phonemic fluency task ranges from .67 to .88 (Baron, 2004).

For the purpose of the present study, the children's total raw scores (i.e., sum of word generated across the three phonemic fluency trials) were standardized to *T* scores using normative data stratified by both age and gender. The choice of gender-stratified norms for the this test followed from data suggesting that, between the ages of nine and 13 years, girls perform significantly better than boys on word fluency (Gaddes & Crockett, 1975). COWAT raw scores for children aged nine to 13 years were normed using data from Gaddes and Crockett (1975), while data from Yeudall, Fromm, Reddon, and Stefanyk (1986) were used for children aged 15 and 16 years. Due to the unavailability of North American normative data for age 14, norms for this age group were interpolated: a mean of 37.6 words (*SD* = 6.85) was used for females and a mean of 32.4 (*SD* = 7.53) was used for males. To obtain the mean for each gender, the total-word raw scores were interpolated by regression and by visual inspection of graphs, and then averaged. Standard deviations for each gender were derived by taking weighted averages of the standard deviations for other age groups in the published normative data.

Trail Making Test (TMT; Reitan, 1958). The TMT is a timed paper-and-pencil test, consisting of two parts. The children's version of the TMT is used for clients aged nine to 14 years, while the adult version is used for clients 15 years of age and older. In Part A (TMT-A), the client is asked to connect numbered circles that are scattered around the page, in numerical order and as quickly as he or she can. The children's version consists of 15 circles, while the adult version includes 25 circles. Part B (TMT-B), which is traditionally thought to be a measure of set shifting, requires the client to connect encircled numbers and letters in alternating order as quickly as he or she can (i.e., 1-A-2-B-3-C-…). Similarly to Part A, the children's version of Part B is comprised of 15 circles, while the adult version of Part B entails 25 circles. Reliabilities for Parts A and B of the Trail Making Test are .98 and .67, respectively (Baron, 2004).

The present study utilizes two scores from the TMT as performance-based measures of EF. The first of these is the traditional time-to-completion score (raw score in seconds) on TMT-B, which was standardized to *T* score form, such that higher *T* scores represented better performance on TMT-B. Age-stratified normative data provided by Knights (1966) were used to standardize scores for children aged nine to 14 years, while data from Fromm-Auch and Yeudall (1983) were used to standardize scores for children aged 15 and 16 years.

The second measure of EF from the TMT that was used in the present study is the TMT B/A ratio (Arbuthnott & Frank, 2000), which is derived by dividing a participant's time (in seconds) to complete TMT-B by his or her time (in seconds) to complete TMT-A. TMT-A, the simpler of the two tasks, constitutes a baseline for motor and visual control and speed. Thus, comparison of TMT-B, the more complex task, to TMT-A reflects the time cost of executive control. According to Lamberty, Putnam, Chatel, Bieliauskas, & Adams (1994), a TMT B/A ratio of three or larger (i.e., taking more than three times longer to complete TMT-B than to complete TMT-A) indicates reduced cognitive efficiency or greater executive dysfunction, while a ratio of less than 2.5 is considered to be normal.

Statistical Analyses

Data were analyzed using the Statistical Package for Social Science (SPSS) for Windows, version 21, and Statistical Analysis System (SAS), version 9.2.

Data Cleaning

Prior to the main analyses, the variables to be used in the present study were checked for accuracy of input of scores through the examination of descriptive statistics. One error of input was suspected based on an implausible value for the GEC score from the BRIEF. The source of the database was contacted and the correct value was obtained.

Next, SPSS Missing Values Analysis was used to identify the amount and pattern of missing data on the demographic variables, estimates of intellectual functioning, and performance-based tests of EF. As shown in Table 3, all variables were missing less than 10% of data. Missing data on the demographic variables (i.e., race, years of education) and estimates of intellectual functioning was not estimated. With respect to neuropsychological test scores, data can be missing for a variety of reasons (e.g., the child is oppositional; he becomes emotionally distressed in response to the task, resulting in discontinuation). In the present study, however, data on the performance-based tests of EF were determined, at least statistically, to be missing at random (Little's MCAR Test: χ^2 = 16.89, p = .815). Accordingly, these data were estimated using expectation maximization.

Table 3

Quantification of missing data on demographic, intellectual functioning, and performance-based EF test variables

Variable	Percent missing
Demographic variables	
Race	2.0
Education	1.0
Intellectual functioning	
Verbal IQ	3.0
Performance IQ	3.0
Full Scale IQ	3.0
Performance-based EF test variables	
WCST perseverative errors	7.1
COWAT total words	8.1
TMT-A time to completion	2.0
TMT-B time to completion	9.1

Notes. WCST = Wisconsin Card Sorting Test; COWAT = Controlled Oral Word Association Test; TMT-A = Trail Making Test Part A; TMT-B = Trail Making Test Part B.

Following the estimation of missing data on the neuropsychological variables, the participants' raw scores for WCST perseverative errors, COWAT total words, and TMT-B time-to-completion were standardized to *T* scores using the normative data identified above. It is notable that some participants performed extremely poorly on the performance-based measures of EF, resulting in negative *T* scores. In light of the facts that (1) *T* scores of zero reflect performance that is already five standard deviations below the mean, and (2) *T* scores in the negative range do not meaningfully alter clinical interpretation beyond this point, all negative *T* scores were replaced with *T* scores of zero. The participants' TMT B/A ratio scores were calculated by dividing time to completion (in seconds) on Part B by time to completion on Part A.

Finally, the data were examined for univariate and multivariate outliers. Based on a leverage cut-off of 0.152, three participants in the original sample of 99 children were identified as outliers on the performance-based tests of EF. Because the removal of these cases influenced the meaningfulness of the results obtained, they were removed from the analyses. Although a standardized residual cut-off of |2.5| resulted in three participants' being identified as outliers on the Behavior Regulation Index (BRI) of the BRIEF, examination of these participants' scores revealed *T* scores that are not uncommon in a clinical setting. Thus, these cases were retained in the analysis. No multivariate outliers were identified as per a Cook's distance cut-off of 1. Removal of three cases that were identified as outliers on the performance-based measures of EF resulted in a final sample of 96 participants in the present study.

Main analyses

As a first step in quantifying the extent of the relationship between performance-based and behavioural measures of EF, alcohol-exposed children's standardized test scores from the WCST, COWAT, and TMT-B were correlated with the eight scales (Inhibit, Shift, Emotional Control, Initiate, Working Memory, Plan/Organize, Organization of Materials, and Monitor) and two indexes (MI and BRI) of the BRIEF. These correlational analyses also served as preliminary analyses for the multiple regression and canonical correlation analyses that followed.

Next, multiple regression analysis was conducted to determine if, and to what extent, scores on the performance-based measures of EF predicted scores on the two indexes of the BRIEF. In both regression analyses, scores from the three performance-based tests of EF were entered as criterion variables in the same step. The Metacognitive Index (MI) of the BRIEF served as the outcome variable in the first analysis, while the Behavior Regulation Index (BRI) comprised the outcome variable in the second analysis.

Finally, for exploratory purposes, a canonical correlation analysis (CCA) was conducted to examine the nature of the relationship between the performance-based EF tests and the BRIEF. CCA is a statistical procedure that identifies and measures the association between two sets of variables by determining a set of canonical variates (Stevens, 2009). As shown in Figure 1, in the CCA relevant to the present study, the two sets of variables are (1) the performance-based EF tests (on the left) and (2) the scales from the BRIEF (on the right). The dark circles represent the canonical variates. Each of the variables within a set contribute some weight in making up the canonical variates on the same side (as represented by the thin, light-coloured lines between the variables and canonical variates). Thus, the column of canonical variates on the left consists of those canonical variates produced by the performance-based EF tests, while the column of canonical variates on the left consists of those canonical variates produced by the scales of the BRIEF. The thicker, dark-coloured lines between the canonical variates on each side represent the correlations (or canonical correlations) between them. The sets of canonical variates (e.g., first canonical variate on the left and first canonical variate on the right, second canonical variate on the left and second canonical variate on the right, and so forth) are generated in such a way (1) so as to maximize the correlation between them, and (2) so that the canonical variates in the second, third, and fourth sets cannot correlate with the variates in the sets above them. Thus, in interpreting a CCA, researchers can not only find out how the variables on each side relate to their canonical variates, but also how strongly the canonical variates from each side relate to each other.

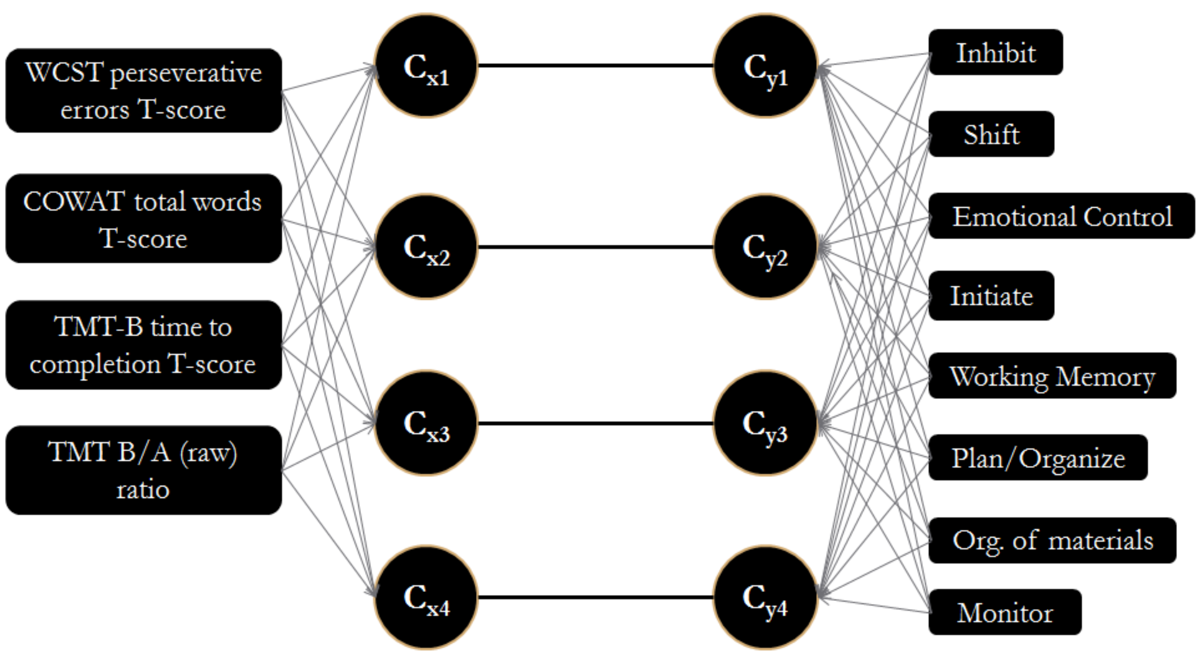

Figure 1. A model of the canonical correlation analysis used to examine the nature of the relationship between performance-based EF test variables and the BRIEF

CHAPTER 3

RESULTS

Assumptions

Prior to conducting the main analyses, all statistical assumptions underlying one or more analyses were evaluated. First, both Pearson product moment correlations and CCA assume interval-level or continuous data (Hair, Anderson, Tatham, & Black, 1998; Tabachnick & Fidell, 2001); consistent with this, all of the variables that were used in the analyses were continuous in nature. Second, all three statistical procedures assume linear relationships between variables (Cohen, Cohen, West, & Aiken, 2003; Tabachnick & Fidell, 2001); visual inspection of bivariate scatterplots indicated that the assumption of linearity was met. Third, univariate normality is an assumption of both Pearson correlations and multiple regression (Cohen et al., 2003), while multivariate normality is an assumption of CCA (Tabachnick & Fidell, 2001). Univariate normality was evaluated using a variety of methods, including a skewness cut-off of |2| and kurtosis cut-off of |4|, Shapiro Wilk's statistic, as well as visual inspection of histograms and normal Q-Q plots (Tabachnick & Fidell, 2001). Taken together, these methods suggested that the assumption of univariate normality was violated for both scores from the TMT (standardized equivalent of the TMT-B time to completion, TMT B/A ratio score), three BRIEF scales (Emotional Control, Working Memory, and Organization of Materials), and one BRIEF index (BRI). Considering that univariate normality is necessary for multivariate normality, the assumption of multivariate normality was taken to be violated. Fourth, Durbin-Watson statistics indicated that the assumption of independence of errors was met for both multiple regression analyses (Durbin-Watson = 1.85 for MRA 1, Durbin-Watson = 1.80 for MRA 2). Next, visual inspection of residual scatterplots suggested that MRA's assumption of homoscedasticity was met (Tabachnick & Fidell, 2001). Finally, although not a formal assumption of the statistical procedures used in the present study, multicollinearity was assessed as it has been shown to affect both MRA and CCA (Cohen et al., 2003; Hair et al., 1998). Based on a Variance Inflation Factor (VIF) cut-off of 10 (Tabachnick & Fidell, 2001), there were no concerns about multicollinearity.

Preliminary Analyses

Descriptive information pertaining to the participants' scores on the performance-based EF tests and the BRIEF is presented in Table 4. With respect to performance-based tests, all scores, with the exception of the TMT B/A ratio, are in *T* score form such that higher scores indicate better performance. While the alcohol-exposed children in our sample performed in the moderately-deficient range on the measure of initiation (COWAT), scores on the two measures traditionally thought to assess shifting were discrepant. On the one hand, the mean score on TMT-B time to completion was moderately-to-severely impaired, and this was corroborated by the mean TMT B/A ratio, which was greater than the cut-off of three provided by Lamberty et al. (1994). On the other hand, their WCST perseverative errors were comparable to neurotypically-developing children of similar age.

On the BRIEF scales and indexes, *T* scores exceeding a value of 65 are considered to be indicative of executive dysfunction (Gioia et al., 2000). As shown in Table 4, alcohol-exposed children in our sample obtained clinically significant scores on all scales and indexes of the BRIEF, with the exception of Organization of Materials.

Table 4

Mean scores of children with FASD on performance-based tests of EF and the BRIEF

Variable	*M*(*SD*)	Range
Performance-based EF tests[a]		
WCST perseverative errors	47.05(9.28)	28.0 – 73.0
COWAT total words	30.47(14.13)	0.00 – 63.0
TMT-B time to completion	23.55(20.10)	0.00 – 66.8
TMT B/A ratio	3.01(1.26)	1.07 – 7.52
BRIEF[b]		
Inhibit	74.47(13.03)	41.0 – 103.0
Shift	73.93(12.27)	40.0 – 95.0
Emotional Control	72.23(13.59)	39.0 – 123.0
Initiate	70.85(10.78)	43.0 – 93.0
Working Memory	75.47(10.17)	45.0 – 93.0
Plan/Organize	72.44(9.15)	53.0 – 103.0
Organization of Materials	62.38(9.91)	34.0 – 98.0
Monitor	71.26(8.10)	47.0 – 91.0
Behavior Regulation Index	76.48(12.15)	44.0 – 109.0
Metacognition Index	74.09(8.35)	54.0 – 92.0

Notes. WCST = Wisconsin Card Sorting Test; COWAT = Controlled Oral Word Association Test; TMT-B = Trail Making Test Part B. TMT = Trail Making Test. BRIEF = Behavior Rating Inventory of Executive Function.
[a]All scores, with the exception of the TMT B/A ratio, are standardized to *T* scores such that higher values reflect better performance (i.e., less impairment). TMT B/A ratios greater than 3.0 denote impairment (Lamberty et al., 1994).
[b]All scores are in the form of *T* scores such that higher scores reflect more impairment. *T* scores greater than 65 are considered to be clinically significant.

Bivariate Correlations

As a first step in quantifying the relationship between the performance-based and behavioural measures of EF, bivariate correlations were run between the four variables from the performance-based EF tests (i.e., standard-score equivalents of WCST perseverative errors, COWAT total words, TMT-B time to completion, and the raw TMT B/A ratio) as well as the scales (Inhibit, Shift, Emotional Control, Initiate, Working Memory, Plan/Organize, Organization of Materials, Monitor) and indexes

(Metacognition Index and Behavior Regulation Index) of the BRIEF. Correlations between scores from the performance-based tests are reported in Table 5 and correlations between the scales and indexes of the BRIEF are reported in Table 6. Of particular interest to the present study, however, are the correlations between scores from the performance-based EF tests and the scales and indexes of the BRIEF. As shown in Table 7, none of these correlations reached statistical significance, with one exception: the standard-score equivalent of the WCST perseverative errors score was significantly correlated with the Organization of Materials scale of the BRIEF ($r = .21$, $p = .037$). In interpreting this correlation, it is important to note that while higher *T* scores on the WCST denote fewer perseverative errors (i.e., better performance), on the BRIEF, higher *T* scores are indicative of greater dysfunction. Thus, although small in magnitude, a positive correlation between the WCST and BRIEF scores suggests that better performance on the WCST is associated with greater impairment on the Organization of Materials scale of the BRIEF.

Table 5

Intercorrelations among performance-based EF test variables

Performance-based EF test score	1	2	3
1. WCST perseverative errors	–		
2. COWAT total words	.23*	–	
3. TMT-B time to completion	.33**	.44**	–
4. TMT B/A ratio	-.26**	.08	-.53**

Notes. All performance-based EF test scores, with the exception of the TMT B/A ratio, were standardized to *T* scores prior to analyses, such that higher scores denoted better performance (i.e., less impairment). WCST = Wisconsin Card Sorting Test; COWAT = Controlled Oral Word Association Test; TMT-B = Trail Making Test Part B; TMT = Trail Making Test.
*p < .05, two-tailed. **p < .01, two-tailed.

Table 6

Intercorrelations among scales and indexes of the BRIEF

BRIEF score	1	2	3	4	5	6	7	8	9
1. Inhibit	–								
2. Shift	.40**	–							
3. Emotional Control	.57**	.59**	–						
4. Initiate	.19	.48**	.18	–					
5. Working Memory	.29**	.48**	.16	.66**	–				
6. Plan/Organize	.09	.28**	.30**	.45**	.50**	–			
7. Organization of Materials	.30**	.25*	.20*	.31**	.44**	.44**	–		
8. Monitor	.33**	.39**	.41**	.53**	.58**	.55**	.28**	–	
9. Behavior Regulation Index	.79**	.73**	.87**	.33**	.34**	.25*	.28**	.46**	–
10. Metacognition Index	.31**	.51**	.33**	.72**	.80**	.75**	.56**	.70**	.38**

Notes. All BRIEF scores were in the form of *T* scores such that higher scores represented greater dysfunction. BRIEF = Behavior Rating Inventory of Executive Function.
* $p < .05$, two-tailed. ** $p < .01$, two-tailed.

Table 7

Correlations between performance-based EF test variables and the scales and indexes of the BRIEF

	Performance-based EF test score[a]			
BRIEF score[b]	WCST perseverative errors	COWAT total words	TMT-B time to completion	TMT B/A ratio
Inhibit	.12	.08	-.08	.06
Shift	.00	-.08	-.17	.003
Emotional Control	.09	-.01	-.09	.004
Initiate	-.02	-.18	-.18	.04
Working Memory	-.06	-.18	-.16	.003
Plan/Organize	.06	-.06	.16	-.14
Organization of Materials	.21*	.11	.07	-.01
Monitor	-.01	-.09	-.03	-.04
Behavior Regulation Index	.05	-.01	-.16	.06
Metacognition Index	.12	-.12	-.04	-.05

Notes. BRIEF = Behavior Rating Inventory of Executive Function; WCST = Wisconsin Card Sorting Test; COWAT = Controlled Oral Word Association Test; TMT-B = Trail Making Test Part B; TMT = Trail Making Test.
[a]All performance-based EF test scores, with the exception of the TMT B/A ratio, were standardized to *T* scores prior to analyses, such that higher scores represented better performance (i.e., less impairment).
[b]All BRIEF scores were in the form of *T* scores, such that higher scores denoted greater dysfunction.
*$p < .05$, two-tailed.

Multiple Regression Analyses

Two multiple regression analyses (MRAs) were used to determine whether scores from performance-based tests of EF could predict Metacognition and Behavior Regulation indexes of the BRIEF, respectively. Given the violation of the assumption of univariate normality for MRA, nonparametric bootstrapping involving 1000 samples was applied to both MRAs (Efron & Tibshirani, 1993).

The first MRA investigated whether the four scores from the performance-based EF tests (i.e., standard-score equivalents of WCST perseverative errors, COWAT total words, TMT-B time to completion, as well as the TMT B/A ratio) predicted scores on the Metacognition Index (MI) of the BRIEF. Because bootstrapping results were not meaningfully discrepant from the results of the original MRA analysis in this case, the latter are reported here. This regression model failed to reach significance (F = .90, p = .469) and accounted for 0% of the variance in MI scores. None of the scores from the performance-based tests emerged as significant predictors of scores on the MI (Table 8).

Table 8

MRA 1: Predictors of Metacognition Index scores

Performance-based test score	*B*	*SE B*	β
Constant	70.17	5.71	
WCST perseverative errors	.15	.10	.165
COWAT total words	-.08	.08	-.136
TMT-B time to completion	-.02	.06	-.038
TMT B/A ratio	-.08	.89	-.012
R^2	.04		

Notes. N = 96. All performance-based EF test scores, with the exception of the TMT B/A ratio, were standardized to T scores prior to analyses such that higher scores represented better performance (i.e., less impairment). WCST = Wisconsin Card Sorting Test; COWAT = Controlled Oral Word Association Test; TMT-B = Trail Making Test Part B; TMT = Trail Making Test.

The second MRA evaluated whether the four scores from the performance-based measures of EF predicted scores on the Behavior Regulation Index (BRI) of the BRIEF. This model was not significant ($F = .94$, $p = .446$), and accounted for 0% of the variance in BRI scores. Scores from the performance-based EF tests did not significantly predict scores on the BRI in the original analysis (Table 9). After bootstrapping, however, the standard-equivalent of the time to completion score on TMT Part B emerged as a significant predictor of the BRI. According to the bootstrap results depicted in Table 10, for every one *T* score increase on the TMT-B, the BRI *T* score decreases by .15 units. These results suggest that better performance on Part B of the TMT (as evidenced by higher *T* scores) very slightly predicts better (i.e., less impaired) scores on the BRIEF (as evidenced by lower *T* scores).

Table 9

MRA 2: Predictors of Behavior Regulation Index scores

Performance-based test score	*B*	*SE B*	β
Constant	73.69	8.31	
WCST perseverative errors	.13	.15	.10
COWAT total words	.07	.11	.08
TMT-B time to completion	-.15	.09	-.25
TMT B/A ratio	-.56	1.30	-.06
R^2	.04		

Notes. $N = 96$. All performance-based EF test scores, with the exception of the TMT B/A ratio, were standardized to *T* scores prior to analyses such that higher scores represented better performance (i.e., less impairment). WCST = Wisconsin Card Sorting Test; COWAT = Controlled Oral Word Association Test; TMT-B = Trail Making Test Part B; TMT = Trail Making Test.

Table 10

MRA 2: Predictors of Behavior Regulation Index scores after bootstrapping with 1000 samples

Performance-based test score	*B*	*SE B*	95% CI
Constant	73.69	8.34	[58.04, 91.15]
WCST perseverative errors	.13	.18	[-.23, .48]
COWAT total words	.07	.10	[-.14, .26]
TMT-B time to completion	-.15*	.07	[-.31, -.00]
TMT B/A ratio	-.56	1.12	[3.18, 1.16]

Notes. $N = 96$. Results obtained after bootstrapping with 1000 samples. All performance-based EF test scores, with the exception of the TMT B/A ratio, were standardized to *T* scores prior to analyses such that higher scores represented better performance (i.e., less impairment). WCST = Wisconsin Card Sorting Test; COWAT = Controlled Oral Word Association Test; TMT-B = Trail Making Test Part B; TMT = Trail Making Test.
*$p < .05$

Canonical Correlation Analysis

In addition to examining the extent of the relationship between performance-based EF tests and the BRIEF, a canonical correlation analysis (CCA) was used to investigate the nature of this relationship, if one existed. Variables in the CCA included the four scores from the performance-based tests (standard-score equivalents of the WCST perseverative error score, COWAT total words score, and TMT-B time to completion score, as well as the TMT B/A ratio) and the eight scales of the BRIEF (Inhibit, Shift, Emotional Control, Initiate, Working Memory, Plan/Organize, Organization of Materials, Monitor). The CCA yielded four canonical correlations. The first pair of canonical variates accounted for 16.2% of the variance; the second for 12.2% of the variance; the third for 2.29% of the variance; and the fourth for 0.10% of the variance.

The results of the significance tests associated with the CCA model that was tested in the present study are depicted in Table 11. It is noteworthy that the significance tests in CCA operate differently than might be expected. Specifically, the first significance test represents an overall test of whether all four canonical correlations together are statistically different from zero; the second test represents an overall test of whether the last three canonical correlations are different from zero; the third test represents an overall test of whether the last two canonical correlations are statistically different from zero; and the fourth test represents an overall test of whether the final canonical correlation is statistically different from zero. As shown in Table 11, none of the tests reached statistical significance. These results suggest that there is no meaningful relationship between scores from the performance-based EF tests and the scales of the BRIEF in our sample.

Table 11

Results of significance testing in canonical correlation analysis

Canonical correlations	Approximate *F*	*p*
1-4	0.94	.563
2-4	0.68	.849
3-4	0.24	.996
4	0.17	.973

Note. N = 96.

CHAPTER 4

DISCUSSION

Executive function (EF) is a higher-order supervisory system that controls or directs the lower-order, domain-specific neuropsychological functions for the purpose of achieving some goal (Gioia & Isquith, 2004). At the very least, it consists of abilities such as planning, initiation, inhibition, cognitive flexibility, and working memory. The multi-domain neuropsychological assessment has long employed a variety of performance-based measures to assess each of these abilities. For example, the Controlled Oral Word Association Test (COWAT) is used widely as a measure of initiation, while the Wisconsin Cart Sorting Test (WCST) and Trail Making Test (TMT) are used to evaluate cognitive flexibility or shifting. Although neuropsychological test performance has formed the basis for inferences about patients' 'real-world' EF capabilities for decades, the ecological validity of these measures has been subjected to empirical investigation only recently, with disappointing results. The present study aimed to advance the literature on the ecological validity of EF tests by examining the extent and nature of the relationship between performance-based and behavioural measures of EF in a relatively large sample of children with Fetal Alcohol Spectrum Disorder (FASD) from Edmonton, Alberta, Canada and surrounding areas. It is noteworthy that nearly 80% of the sample in this study consisted of children of Aboriginal descent, and this is consistent with some studies that have found higher rates of FASD in Aboriginal populations in Canada (Asante & Nelms-Mazike, 1985; Muckle et al., 2005; Robinson et al., 1987; Square, 1997).

Performance on measures of EF

As part of the present study, the performance of children with FASD was evaluated on performance-based and behavioural measures of EF. While their impaired scores on the COWAT and TMT were consistent with past studies documenting deficits in phonemic fluency (e.g., Aragon et al., 2008) and set shifting (e.g., Vaurio et al., 2008), respectively, among this population, their commission of perseverative errors on the WCST, another indicator of set-shifting ability, was comparable to neurotypically-developing children of similar age. The latter finding was inconsistent with some past studies in which alcohol-exposed children made more perseverative errors on card sorting tasks than controls (Kodituwakku et al., 1995, 2001; Vaurio et al., 2008). However, in interpreting this finding, it is important to remember that the perseverative-errors score is only one of many indicators of performance on the WCST. Indeed, studies examining other scores from the WCST (e.g., total errors, non-perseverative errors) suggest that children with FASD tend to respond in a disorganized, unplanned manner on this test (e.g., Olson et al., 1998; Vaurio et al., 2008). Although such a pattern of responding could not be confirmed in our sample due to the unavailability of total and non-perseverative error scores, it was determined that only 35% of the children in the sample successfully completed six categories.

Based on caregiver ratings, alcohol-exposed children in our sample obtained clinically-significant scores on all scales and indexes of the BRIEF, with the exception of Organization of Materials, which was in the normal range. This pattern of scores is largely consistent with the findings of Rasmussen and colleagues (2007). It is noteworthy, however, that while Organization of Materials was an absolute strength for the alcohol-exposed children in our study (i.e., the mean score was in the normal range), this scale represented a relative strength for the children in their study (i.e., their mean score was clinically significant). It has been suggested that the finding that children with FASD struggle the least with Organization of Materials may be confounded by their caregivers' tendency to maintain highly structured living and play areas (Rasmussen et al., 2007).

Relationship between performance-based EF tests and BRIEF

As a first step in quantifying the relationship between the performance-based tests of EF (i.e., WCST, COWAT, and TMT) and the BRIEF, bivariate correlations between the scores from these measures were examined. While small but significant correlations had been expected, this hypothesis was not supported. In fact, the correlations between the EF tests and the scales and indexes of the BRIEF failed to reach statistical significance in all but one case. Only the standard-equivalent of the perseverative errors score from the WCST was significantly correlated with the Organization of Materials scale of the BRIEF, such that better performance (i.e., fewer perseverative errors) on the WCST was associated with greater impairment on the Organization of Materials scale. Although not in the expected direction, this correlation is not implausible. The Organization of Materials scale of the BRIEF assesses a child's ability to organize and keep track of his environment and possessions (Gioia et al., 2000). With items such as "Keeps room messy", "Cannot find things in room or school desk", and "Leaves a trail of belongings wherever he/she goes", this scale characterizes a child who engages with objects in his or her world in a highly unsystematic manner. In contrast to this, the commission of perseverative errors on the WCST in particular (i.e., as opposed to non-perseverative errors), requires a very systematic approach. The WCST perseverative error score is a measure of how often a child attempts to match the stimulus card to the key card on the basis of a principle that is no longer correct. In order to make perseverative errors, then, a child must keep track of (1) his most recent match between stimulus card and key card that produced a correct answer and (2) which of the three matching criteria (i.e., colour, form, or number) was operating during that sort. Once this matching criterion is identified, he must hold it in mind and match new or upcoming stimulus cards in accordance with it, while inhibiting distraction by other undesired matching criteria. Considering this, it is reasonable to understand that a child who has difficulty maintaining orderliness with respect to his work, play, and storage spaces in the real world might obtain a low WCST perseverative errors score simply because he cannot keep track of all of the information that is required for him to commit perseverative errors. Instead, such a child is probably more likely to make an increased number of non-perseverative errors on the WCST as a result of matching stimulus cards to key cards randomly. As mentioned above, the latter possibility could not be tested empirically in the present study as total and non-perseverative error scores from the WCST were not available in the archival dataset that was used.

In addition to the examination of bivariate correlations, the relationship between the performance-based and behavioural measures of EF was further quantified by two multiple regressions in which scores from the performance-based tests were used to predict index scores from the BRIEF. Scores from the performance-based tests failed to predict scores on the Metacognition Index of the BRIEF. However, Part B of the Trail Making Test (TMT-B) emerged as a significant predictor of the Behavior Regulation

Index after bootstrapping with 1000 samples. Although significant, the meaningfulness of this finding is suspect. First, as shown in Table 7, the standard-score equivalent of the time-to-completion score on the TMT-B does not correlate significantly with the BRI. Second, according to Shear and Zumbo (2013), Type I error rates can become inflated when (1) the predictor variables are correlated with each other and (2) one or more of the predictor variables contains random measurement error. As shown in Table 5, all but one of the correlations between the predictor variables were found to be statistically significant. Furthermore, the reliability coefficients associated with the performance-based measures indicate that all of the predictor variables in the regression model contained random measurement error. Thus, it appears that the significance of TMT-B as a predictor of the BRI constitutes a false positive (i.e., a Type I error).

Finally, in an effort to examine the nature of the relationship between the performance-based measures and the BRIEF, a canonical correlation analysis (CCA) was conducted. The CCA yielded four canonical correlations, none of which reached statistical significance.

Although not statistically significant, it was noted that the variates comprising the first canonical correlation accounted for 16.2% of variance. Considering that this represents nearly one-fifth of the total variance, the first canonical correlation was interpreted. The first X canonical variate was found to be dominated by the children's standardized TMT-B time-to-completion scores, while the first Y canonical variate was dominated by both the Working Memory and Plan/Organize scales of the BRIEF. These results provide preliminary evidence for a relationship between time-to-completion on Part B of the Trail Making Test and the Working Memory and Plan/Organize scales of the BRIEF, which may not have reached statistical significance in the current study due to insufficient power.

Taken together, the results of this study fail to provide evidence for a meaningful relationship between performance-based EF tests and the BRIEF. These findings are consistent with other studies that have found poor ecological validity of performance-based tests of EF in various clinical samples of children (Conklin et al., 2008; MacAllister et al., 2012; Vriezen & Pigott, 2002). Even in studies with significant findings, single EF tests have accounted for no more than 10% of shared variance, while optimal combinations of EF tests accounted for 12-20%.

In understanding the small, if practically non-existent, relationship between performance-based EF tests and the BRIEF, several issues must be considered. The first of these is the use of the BRIEF in studies examining the ecological validity of traditional, performance-based tests of EF in children. The use of the BRIEF as a comparison standard for performance-based EF measures assumes that the BRIEF itself has high, if not optimal, ecological validity. If we define ecological validity in terms of verisimilitude and veridicality (Chaytor & Schmitter-Edgecombe, 2003; Franzen & Wilhelm, 1996), the BRIEF appears to possess impressive ecological validity. With respect to verisimilitude, although the respondent (parent, teacher, or child) may be completing the rating scale in a quiet workspace that is free from distractions, the items on the scale require answers based on the child's behaviour in his or her home or school, environments that are often far from quiet and full of distractions. With respect to veridicality, the BRIEF has been shown to be related to other measures of impairment, such as measures of attentional, behavioural, and socio-emotional problems (McAuley, Chen, Goos, Schachar, & Crosbie, 2010) and adaptive functioning (Gilotty, Kenworthy, Sirian, Black, & Wagner, 2002).

If we accept the BRIEF as an ecologically valid measure of EF, the absence of meaningful relationships between performance-based EF tests and the BRIEF demonstrates that EF tests fare poorly in terms of veridicality. Considered alongside the barriers to verisimilitude of EF tests discussed earlier (e.g., overly simplistic and novel tasks, artificial testing environment), these results can be taken as evidence for the idea that performance-based tests of EF have little, if any, ecological validity when used with alcohol-exposed children.

In the literature, the small, and sometimes non-existent, association between performance-based and behavioural measures of EF has been explained with the ideas that (1) the two types of measures evaluate different aspects of EF (Barkley & Murphy, 2010), or (2) that performance-based tasks assess underlying skills while ratings evaluate the application of those skills in real-world settings, such as the home or school (McAuley et al., 2010). Unfortunately, however, there has been little speculation with respect to the nature of these differences. The present discussion aims to grapple with this issue.

The early theoretical distinction between metacognitive and emotional/motivational EF was based on brain-behaviour relationships: Metacognitive EF were subserved by the dorsolateral prefrontal cortex, whereas emotional or motivational EF were subserved by the ventromedial prefrontal cortex (Cummings, 1993). When the dorsolateral prefrontal cortex was damaged, impaired metacognitive EF affected a patient’s cognition or cognitive control, leaving him unable to organize his behavioural responses to novel or complex stimuli, solve problems, and attain external goals. On the other hand, when the ventromedial prefrontal cortex was damaged, impaired emotional/motivational EF influenced a patient's behaviour: as a result of the inability to fulfill his basic impulses in a socially acceptable way, he was irritable, tactless, distractible, and often found engaging in inappropriate behaviours.

In accordance with this seemingly clear-cut distinction between metacognitive and emotional/motivational EF, the traditional performance-based tests of EF were employed to evaluate individual metacognitive EF abilities in neuropsychological assessment. Specifically, the tower tasks were thought to assess planning; fluency tests to evaluate initiation; the Stroop task to emphasize inhibition; the WCST and TMT to measure cognitive flexibility; and the Digit Span task to test working memory.

Despite its usefulness for the conceptualization and assessment of EF in neuropsychology, the theoretical distinction between metacognitive and emotional/motivational EF appears to be irrelevant in the real world. Indeed, the vast majority of problems that individuals face on a daily basis (e.g., deciding how to approach a friend, co-worker, or boss; spending or investing money) are not emotionally neutral (Ardila, 2008). This point is illustrated by the composition of the BRIEF, a behavioural measure of EF which is considered to have at least acceptable ecological validity.

At first glance, the BRIEF's two indexes, the Metacognition Index (MI) and the Behavior Regulation Index (BRI), suggest that the measure adheres to the distinction between metacognitive and emotional/motivational EF. However, the factor structure of the BRIEF suggests otherwise. First, while many EF theorists agree that planning, initiation, inhibition, cognitive flexibility, and working memory are fundamental subcomponents of metacognitive EF, on the BRIEF, the Inhibit and Shift scales load on the BRI. The contribution of the Inhibit and Shift scales to the BRI, as opposed to the MI, indicates that these subcomponents of EF are not emotionally neutral. An examination of the items comprising these scales reveals statements involving a social and/or emotional context. For example, the Inhibit scale contains items such as “Interrupts others”, “Acts wilder or sillier than others in groups (birthdays, recess)”, while the Shift scale includes items such as “Resists accepting a different way to solve a problem with schoolwork, friends, chores, and so on” and “Acts upset by a change in plans”.

Second, even for items that load on the MI, the absence of emotional content in the items themselves does not preclude the possibility that emotional/motivational factors can influence the child’s behaviour and, by extension, the informant’s response to the item(s). Consider, for example, the following items: “Has trouble taking action to reach goals (saving money for special item; studying to get good grades)” (Plan-Organize), “Has trouble getting started on homework or chores” (Initiation), and “Has trouble concentrating on chores, schoolwork” (Working Memory). Although activities such as studying and completing homework are beneficial to children in the long term, they are not immediately gratifying: it is not difficult to fathom that most children, even those who enjoy and do well in school, would rather be playing a videogame or a sport, watching a movie, or spending time with their friends than studying or doing chores. Thus, although a respondent’s answer of “sometimes” or “often” to the item “Has trouble concentrating on chores, schoolwork” might indicate that the child genuinely has metacognitive executive dysfunction, it could also reflect a motivational issue: the child may have difficulty concentrating on his schoolwork simply because he is more interested in an immediately-gratifying activity such as eating, watching a show on TV, or playing a videogame with his sibling. Because the BRIEF does not account for the latter possibility, the MI cannot be considered to be free of an emotional/motivational component.

The arguments presented above illustrate two important points about the BRIEF in particular and about real-world manifestation of EF more generally. First, although the names of the BRIEF indexes suggest that they differentiate between metacognitive and emotional/motivational EF, each index can be, and likely is, influenced by both metacognitive and emotional/motivational aspects of EF. Thus, in the real world, metacognitive and emotional/motivational EF are intrinsically linked. Second, while both metacognitive and emotional/motivational EF are critical to an individual's successful functioning in the real world, appropriate behavioural regulation, which relies largely on emotional/motivational EF, is likely a precursor for metacognitive problem solving (Gioia et al., 2000): Only a child who can successfully resist the impulse to engage in more immediately-gratifying activities is likely to complete his homework or chores.

The above discussion highlights the relationship between metacognitive and emotional/motivational EF (1) as it manifests in the real world, and (2) as it is assessed by the BRIEF. Performance-based EF tests, on the other hand, attempt to evaluate isolated metacognitive EF processes in an emotionally-neutral context. Considering this discrepancy, it is not surprising that studies examining the ecological validity of performance-based EF tests have failed to find a meaningful relationship between the two types of measures.

Of course, it can be argued that the use of the BRIEF as a comparison standard for evaluating the ecological validity of performance-based EF tests is unfair since the two types of measures do not assess the same constructs. However, the stronger argument is probably the opposite: in fact, it is unfair to strip emotional/motivational aspects of executive function from EF tests and then use these tests to make inferences about the patient's behaviour in a context that constantly requires the integration of both metacognitive and emotional/motivational EF.

Limitations

The results of the present investigation should be interpreted in light of the study's limitations. The first of these limitations concerns the normative data used in the study: because the use of retrospective data precluded the inclusion of a control group of neurotypically-developing children, the participants' scores on both the performance-based EF tests and the BRIEF were standardized according to existing normative data. Whereas nearly 80% of the children included in our sample were Aboriginal, the data that was used to norm their scores on measures of EF was based largely on Caucasian children. Given the many cultural and environmental differences between Aboriginal and non-Aboriginal children, it is clear that the normative data used in this study were less than optimal. In addition to being based primarily on Caucasian children, normative data for the performance-based EF tests, in particular, were not only significantly dated, but in the case of the COWAT, also incomplete: COWAT norms for were interpolated for 10 males and 5 females aged 14 years. These limitations highlight not only the need for the publication of normative data specific to Aboriginal children, but also the need for more current and more complete normative data sets, in general, for older neuropsychological tests that remain in use today.

Second, despite the fact that the current study involved a larger sample size than some other studies examining the ecological validity of EF tests, it may have lacked sufficient power to detect an existing effect. More specifically, the final sample used for the purpose of the analyses consisted of 96 children on the fetal alcohol spectrum. Although a sample of this size is sufficient to find a Pearson correlation of medium effect, it is not large enough to detect a correlation of small effect (Cohen, 1992). Likewise, according to guidelines provided by Tabachnick and Fidell (2001), the sample was slightly smaller in size than what is required to find a medium effect (i.e., 107 cases).

Finally, qualitative information pertaining to the level of effort or engagement of each participant with the performance-based EF tasks (i.e., WCST, COWAT, TMT) in particular was unavailable. Thus, although it was assumed that all of the participants put forth a full effort on the performance-based EF tests, this may not have been the case.

Directions for Future Research

The present study was the first, to our knowledge, to investigate the relationship between performance-based EF tests and the BRIEF in a sample of children with Fetal Alcohol Spectrum Disorder (FASD). While this study makes an important contribution to research examining the ecological validity of performance-based tests of EF, it also highlights several directions for future research. First, considering the insufficient power of the present study to detect effects of small magnitude, it may be worth replicating this study in a larger sample. Second, the present study grouped children at various points along the fetal alcohol spectrum into a single group. Perhaps investigations of the ecological validity of performance-based EF tests may yield different results when examined in children with FAS, pFAS, ARND, and ARBD separately. Third, whereas the present study only included three performance-based measures of EF (i.e., WCST, COWAT, TMT), future research could examine the ecological validity of other traditional performance-based EF tests, such as the Stroop or Digit Span tasks, in children with FASD. Furthermore, instead of relying exclusively on parent or caregiver ratings, the inclusion of ratings from multiple informants (e.g., teachers) may be worth investigating.

In light of the increasing number of studies demonstrating the low ecological validity of EF tests, the last two decades have witnessed the development of performance-based tasks of EF that are characterized by higher verisimilitude than the traditional EF tests. Although some studies involving adult populations suggest that tasks with higher verisimilitude have superior veridicality than traditional performance-based EF tests (Chaytor & Schmitter-Edgecombe, 2003), future research should examine the ecological validity of high-verisimilitude executive tests for children, such as the Modified Six Elements Test-Children's version (Siklos & Kerns, 2004) and the Behavioural Assessment of the Dysexecutive Syndrome in Children (Emslie, Wilson, Burden, Nimmo-Smith, & Wilson, 2003).

Finally, if the goal of neuropsychological assessment is to predict children's real-world functioning, ecological validity of neuropsychological measures is paramount. Considering this, it may be worthwhile to re-examine the traditional performance-based measures as measures of other neuropsychological domains (e.g., attention). It is possible that these measures possess high ecological validity in other domains.

Conclusions

The present study investigated the ecological validity of performance-based tests of EF in a sample of children with Fetal Alcohol Spectrum Disorder (FASD). To do so, scores from three performance-based EF tests were compared to caregiver ratings on the BRIEF, a popular behaviour rating scale of EF. Results failed to provide evidence for a meaningful relationship between the performance-based and behavioural measures. Although disappointing, the findings of this study are in keeping with the literature in the area, which suggests that performance-based EF tests have low, if any, ecological validity in various clinical populations of children. This trend of findings has been explained by the idea that performance-based EF tests and behavioural measures of EF assess different aspects of EF. The present paper attempts to explore these differences. Specifically, we argue that while performance-based tests assess individual metacognitive functions (i.e., planning, initiation, inhibition, cognitive flexibility, working memory), in the real world, metacognitive and emotional/motivational aspects of EF are intrinsically linked. Finally, limitations of the present study and directions for future research are identified.

REFERENCES

Aase, J. M. (1981). The fetal alcohol syndrome in American Indians: a high risk group. *Neurobehav Toxicol Teratol, 3*(2), 153-156.

Abel, E. L. (1995). An update on incidence of FAS: FAS is not an equal opportunity birth defect. *Neurotoxicol Teratol, 17*(4), 437-443.

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

Anderson, S. W., Damasio, H., Dallas, J. R., & Tranel, D. (1991). Wisconsin card sorting test performance as a measure of frontal lobe damage. *Journal of Clinical and Experimental Neuropsychology, 13*(6), 909-922. doi: 10.1080/01688639108405107

Anderson, S. W., Bechara, A., Damasio, H., Tranel, D. & Damasio, A. R. (1999). Impairment of social and moral behavior related to early damage in human prefrontal cortex. *Nature Neuroscience, 2*(11), 1032-1037.

Anderson, V. A., Anderson, P., Northam, E., Jacobs, R., & Mikiewicz, O. (2002). Relationships between cognitive and behavioral measures of executive function in children with brain disease. *Child Neuropsychology, 8*(4), 231-240. doi: 10.1076/chin.8.4.231.13509

Aragon, A. S., Kalberg, W. O., Buckley, D., Barela-Scott, L. M., Tabachnick, B. G., & May, P. A. (2008). Neuropsychological study of FASD in a sample of American Indian children: processing simple versus complex information. *Alcohol Clin Exp Res, 32*(12), 2136-2148. doi: 10.1111/j.1530-0277.2008.00802.x

Arbuthnott, K., & Frank, J. (2000). Trail Making Test, Part B as a measure of executive control: validation using a set-switching paradigm. *Journal of Clinical & Experimental Neuropsychology, 22*(4), 518-528.

Ardila, A. (2008). On the evolutionary origins of executive functions. *Brain and Cognition, 68*(1), 92-99. doi: http://dx.doi.org/10.1016/j.bandc.2008.03.003

Army Individual Test Battery. (1944). *Manual of directions and scoring*. Washington DC: War Department, Adjutant General's office.

Asante, K., & Nelms-Mazike, J. (1985). *Report on the survey of children with chronic handicaps and fetal alcohol syndrome in the Yukon and Northwest British Columbia*. Whitehorse, YT: Council for Yukon Indians.

Astley, S. J., Olson, H. C., Kerns, K., Brooks, A., Aylward, E. H., Coggins, T. E., . . . Richards, T. (2009). Neuropsychological and behavioral outcomes from a comprehensive magnetic resonance study of children with fetal alcohol spectrum disorders. *Canadian Journal of Clinical Pharmacology, 16*(1), e178-e201.

Bakar, E., Taner, Y., Soysal, A., Karakas, S., & Turgay, A. (2011). Behavioral Rating Inventory and Laboratory Tests Measure Different Aspects of Executive Functioning in Boys: A Validity Study. *Bulletin of Clinical Psychopharmacology*, 1. doi: 10.5455/bcp.20111004014003

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. *Arch Clin Neuropsychol, 25*(3), 157-173. doi: 10.1093/arclin/acq014

Baron, I. S. (2004). *Neuropsychological Evaluation of the Child*. New York, NY: Oxford University Press.

Barr, H. M. & Streissguth, A. P. (2001). Identifying maternal self-reported alcohol use associated with fetal alcohol spectrum disorders. *Alcoholism: Clinical and Experimental Reserarch, 25*(2), 283-287.

Barrash, J., Tranel, D., & Anderson, S. W. (2000). Acquired personality disturbances associated with bilateral damage to the ventromedial prefrontal region. *Developmental Neuropsychology, 18*(3), 355-381.

Benton, A. L., & Hamsher, K. (1976). *Multilingual aphasia examination*. Iowa City: University of Iowa.

Berg, E. A., & Grant, D. A. (1948). The performance of topectomized patients on the University of Wisconsin Card Sorting Test. *American Psychologist, 3*, 360.

Bialystok, E., & Martin, M. M. (2004). Attention and inhibition in bilingual children: evidence from the dimensional change card sort task. *Developmental Science, 7*(3), 325-339. doi: 10.1111/j.1467-7687.2004.00351.x

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

Bray, D. L. & Anderson, P. D. (1989). Appraisal of the epidemiology of fetal alcohol syndrome among Canadian Native peoples. *Canadian Journal of Public Health, 80*(1), 42-45.

Brunswik, E. (1955). Symposium on the probability approach in psychology. *Psychological Review, 62*(3), 193-217.

Centers for Disease Control and Prevention (2002). FAS-Alaska, Arizona, Colorado, and New York 1995-1997. *Morbidity and Mortality Weekly Report, 514*, 433-435.

Chasnoff, I. J., Wells, A. M., Telford, E., Schmidt, C., & Messer, G. (2010). Neurodevelopmental functioning in children with FAS, pFAS, and ARND. *Journal of Development and Behavioral Pediatrics, 31*(3), 192-201. doi: 10.1097/DBP.0b013e3181d5a4e2

Chaytor, N., & Schmitter-Edgecombe, M. (2003). The ecological validity of neuropsychological tests: A review of the literature on everyday cognitive skills. *Neuropsychology Review, 13*(4), 181-197. doi: 10.1023/B:NERV.0000009483.91468.fb

Chaytor, N., Schmitter-Edgecombe, M., & Burr, R. (2006). Improving the ecological validity of executive functioning assessment. *Arch Clin Neuropsychol, 21*(3), 217-227. doi: 10.1016/j.acn.2005.12.002

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155-159.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regressionl correlation analysis for the behavioral sciences* (3rd ed.). Hillsdale, NJ: Erlbaum.

Coles, C. D., Platzman, K. A., Raskind-Hood, C. L., Brown, R. T., Falek, A., & Smith, I. E. (1997). A comparison of children affected by prenatal alcohol exposure and attention deficit, hyperactivity disorder. *Alcoholism: Clinical & Experimental Research, 21*(1), 150-161. doi: 10.1111/j.1530-0277.1997.tb03743.x

Conklin, H. M., Salorio, C. F., & Slomine, B. S. (2008). Working memory performance following paediatric traumatic brain injury. *Brain Inj, 22*(11), 847-857. doi: 10.1080/02699050802403565

Cummings, J. L. (1993). Frontal-subcortical circuits and human behavior. *Archives of Neurology, 50*, 873-880. doi: 10.1001/archneur.1993.00540080076020

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

Donders, J. (2002). The behavior rating inventory of executive function: Introduction. *Child Neuropsychology, 8*(4), 229-230. doi: 10.1076/chin.8.4.229.13508

Efron, B., & Tibshirani, R. (1993). *An introduction to the bootstrap*. New York: Chapman and Hall.

Emslie, H., Wilson, F. C., Burden, V., Nimmo-Smith, I., & Wilson, B. A. (2003). *Behavioural assessment of the dysexecutive syndrome in children (BADS-C).* London, U.K.: Harcourt Assessment.

Eslinger, P.J. & Damasio, A.R. (1985). Severe disturbance of higher cognition after bilateral frontal lobe ablation: patient EVR. *Neurology*, *35*(12), 1731-1741.

Eslinger, P.J., Grattan, L.M., & Damasio, A.R. (1992). Developmental consequences of childhood frontal lobe damage. *Arch Neurol*, *49*, 764-769.

Espy, K. A., Stalets, M. M., McDiarmid, M. M., Senn, T. E., Cwik, M. F., & Hamby, A. (2002). Executive functions in preschool children born preterm: Application of cognitive neuroscience paradigms. *Child Neuropsychology, 8*(3), 83-92. doi: 10.1076/chin.8.2.83.8723

Franzen, M. D., & Wilhelm, K. L. (1996). conceptual foundations of ecological validity in neuropsychological assessment. In R. J. Sbordone & J. Long (Eds.), *Ecological validity of neuropsychological testing*. Boca Raton, FL: St. Lucie.

Fromm-Auch, D., & Yeudall, L. T. (1983). Normative data for the Halstead-Reitan Neuropsychological Tests. *Journal of Clinical neuropsychology, 5*, 221-238.

Fuster, J. M. (2001). The prefrontal cortex-An update: Time is of the essence. *Neuron, 30*, 319-333. doi: 10.1016/S0896-6273(01)00285-9

Fuster, J. M. (2002). Frontal lobe and cognitive development. *Journal of Neurocytology, 31*(373-385). doi: 10.1023/A:1024190429920

Gaddes, W., & Crockett, D. (1975). The Spreen-Benton aphasia tests: Normative data as a measure of normal language development. *Brain and Language, 2*, 257-280.

Gilotty, Lisa, Kenworthy, Lauren, Sirian, Lisa, Black, David O., & Wagner, Ann E. (2002). Adaptive skills and executive function in autism spectrum disorders. *Child Neuropsychology*, *8*(4), 241-248. doi: 10.1076/chin.8.4.241.13504

Gioia, G. A., & Isquith, P. K. (2004). Ecological assessment of executive function in traumatic brain injury. *Developmental Neuropsychology, 25*(1&2), 135-158. doi: 10.1207/s15326942dn2501&2_8

Gioia, G. A., Isquith, P. K., Guy, S. C., & Kenworthy, L. (2000). Behavior rating inventory of executive function. *Child Neuropsychology*, *6*(3), 235-238. doi: 10.1076/chin.6.3.235.3152

Gioia, G. A., Isquith, P. K., Guy, S. C., & Kenworthy, L. (2000). *BRIEF: Behavior rating inventory of executive function*. Lutz, FL: Psychological Assessment Resources Inc.

Gioia, G. A., Isquith, P. K., & Kenworthy, L. (2000). *BRIEF - Behavior Rating Inventory of Executive Function, Professional Manual*. Odessa, FL: Psychological Assessment Resources.

Gioia, G. A., Isquith, P. K., Kenworthy, L., & Barton, R. M. (2002). Profiles of everyday executive function in acquired and developmental disorders. *Child Neuropsychology, 8*(2), 121-137. doi: 10.1076/chin.8.2.121.8727

Gioia, G. A., Isquith, P. K., Retzlaff, P. D., & Espy, K. A. (2002). Confirmatory factor analysis of the behaviour rating inventory of executive function (BRIEF) in a clinical sample. *Child Neuropsychology, 8*(4), 249-257. doi: 10.1076/chin.8.4.249.13513

Gioia, G. A., Isquith, P. K., Retzlaff, P. D., & Pratt, B.M. (2001). Modeling executive functions with everyday behaviors: A unitary or fractioned system? *Brain and Cognition, 45*.

Green, C. R., Mihic, A. M., Nikkel, S. M., Stade, B. C., Rasmussen, C., Munoz, D. P., & Reynolds, J. N. (2009). Executive function deficits in children with fetal alcohol spectrum disorders (FASD) measured using the Cambridge Neuropsychological Tests Automated Battery (CANTAB). *J Child Psychol Psychiatry, 50*(6), 688-697. doi: 10.1111/j.1469-7610.2008.01990.x

Griffith, E. M., Pennington, B. F., Wehner, E. A., & Rogers, S. J. (1999). Executive functions in young children with autism. *Child Development, 70*(4), 817-832. doi: 10.1111/1467-8624.00059

Hair, J. F. Jr., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate Data Analysis* (5th ed.). Upper Saddle River, NJ: Prentice Hall, Inc.

Happaney, K., Zelazo, P. D., & Stuss, D. T. (2004). Development of orbitofrontal function: current themes and future directions. *Brain Cogn, 55*(1), 1-10. doi: 10.1016/j.bandc.2004.01.001

Heaton, R. K., Chelune, G. J., Talley, J. L., Kay, G. G., & Curtiss, G. (1993). *Wisconsin card sorting test manual - revised and expanded*. Odessa, FL: Psychological Assessment Resources, Inc.

Hill, E. L. (2004). Evaluating the theory of executive dysfunction in autism. *Developmental Review, 24*(2), 189-233. doi: 10.1016/j.dr.2004.01.001

Hoyme, H. E., May, P. A., Kalberg, W. O., Kodituwakku, P. W., Gossage, J. P., Trujillo, P. M., . . . Robinson, L. (2005). A practical clinical approach to diagnosis of fetal alcohol spectrum disorders: clarification of the 1996 Institute of Medicine criteria. *Pediatrics, 115*(1), 39-47.

Kaplan, E., Fein, D., Kramer, J., Delis, D., & Morris, R. (1999). *WISC-III-PI manual*. San Antonio, TX: The Psychological Corporation.

Kimberg, D. Y. (1996). Executive functions, working memory, and frontal lobe function. *Dissertation Abstracts International: Section B: The Sciences & Engineering, 57*(1-B), 0736.

Knights, R. M. (1966). *Normative data on tests evaluating brain damage in children 5-14 years of age*. London, ON: Department of Psychology, University of Western Ontario.

Kodituwakku, P. W. (1993). *Progressive Planning Test: Manual*. Albuquerque, NM: University of New Mexico Department of Psychology.

Kodituwakku, P. W., Handmaker, N. S., Cutler, S. K., Weathersby, E. K., & Handmaker, S. D. (1995). Specific impairments in self-regulation in children exposed to alcohol prenatally. *Alcoholism: Clinical & Experimental Research, 19*(6), 1558-1564. doi: 10.1111/j.1530-0277.1995.tb01024.x

Kodituwakku, P. W., May, P. A., Clericuzio, Carol L., & Weers, David. (2001). Emotion-related learning in individuals prenatally exposed to alcohol: an investigation of the relation between set shifting, extinction of responses, and behavior. *Neuropsychologia, 39,* 699-708.

Krikorian, R., & Bartok, J. A. (1998). Developmental Data for the Porteus Maze Test. *The Clinical Neuropsychologist (Neuropsychology, Development and Cognition: Section D), 12*(3), 305-310. doi: 10.1076/clin.12.3.305.1984

Lamberty, G. J., Putnam, S. H., Chatel, D. M., Bieliauskas, L. A., & Adams, K. M. (1994). Derived Trail Making Test indices. *Neuropsychiatry, Neuropsychology, and Behavioral Neurology, 7,* 230-234.

Levin, H. S., Hanten, G., Zhang, L., Swank, P. R., & Hunter, J. (2004). Selective impairment of inhibition after TBI in children. *Journal of Clinical and Experimental Neuropsychology, 26*(5), 589-597. doi: 10.1080/13803390490504434

Lezak, M. D., Howieson, D. B., Loring, D. W., Hannay, H. J., & Discher, J. S. (2004). *Neuropsychological assessment* (4 ed.). London: Oxford University Press.

MacAllister, W. S., Bender, H. A., Whitman, L., Welsh, A., Keller, S., Granader, Y., & Sherman, E. M. (2012). Assessment of executive functioning in childhood epilepsy: the Tower of London and BRIEF. *Child Neuropsychol, 18*(4), 404-415. doi: 10.1080/09297049.2011.613812

MacMillan, H. L., Jamieson, E., Walsh, C. A., Wong, M. Y.-Y., Faries, E. J., McCue, H., MacMillan, A. B., Offord, D. R., & The Technical Advisory Committee of the

Chiefs of Ontario. First Nations women's mental health: results from an Ontario survey. *Arch Womens Ment Health, 11*, 109-115.

Maillard-Wermelinger, A., Yeates, K. O., Taylor, H. G., Rusin, J., Bangert, B., Dietrich, A., . . . Wright, M. (2009). Mild traumatic brain injury and executive functions in school-aged children. *Developmental Neurorehabilitation, 12*(5), 330-341. doi: 10.1080/17518420903087251

Mattson, S. N., Crocker, N., & Nguyen, T. T. (2011). Fetal alcohol spectrum disorders: neuropsychological and behavioral features. *Neuropsychol Rev, 21*(2), 81-101. doi: 10.1007/s11065-011-9167-9

Mattson, S. N., Goodman, A. M., Caine, C., Delis, D. C., & Riley, E. P. (1999). Executive functioning in children with heavy prenatal alcohol exposure. *Alcoholism: Clinical & Experimental Research, 23*(11), 1808-1815. doi: 10.1097/00000374-199911000-00014

Mattson, S. N., & Riley, E. P. (1998). A review of the neurobehavioral deficits in children with fetal alcohol syndrome or prenatal exposure to alcohol. *Alcoholism: Clinical and Experimental Research, 22*(2), 279-294.

Mattson, S. N., Riley, E. P., Gramling, L., Delis, D. C., & Jones, K. L. (1998). Neuropsychological comparison of alcohol-exposed children with or without physical features of fetal alcohol syndrome. *Neuropsychology, 12*(1), 146-153.

Mattson, S. N., Roesch, S. C., Fagerlund, A., Autti-Ramo, I., Jones, K. L., May, P. A., . . . Collaborative Initiative on Fetal Alcohol Spectrum Disorders. (2010). Toward a neurobehavioral profile of fetal alcohol spectrum disorders. *Alcohol Clin Exp Res, 34*(9), 1640-1650. doi: 10.1111/j.1530-0277.2010.01250.x

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? *J Int Neuropsychol Soc, 16*(3), 495-505. doi: 10.1017/S1355617710000093

McGee, C. L., Fryer, S. L., Bjorkquist, O. A., Mattson, S. N., & Riley, E. P. (2008). Deficits in social problem solving in adolescents with prenatal exposure to alcohol. *Am J Drug Alcohol Abuse, 34*(4), 423-431. doi: 10.1080/00952990802122630

McGee, C. L., Schonfeld, A. M., Roebuck-Spencer, T. M., Riley, E. P., & Mattson, S. N. (2008). Children with heavy prenatal alcohol exposure demonstrate deficits on multiple measures of concept formation. *Alcohol Clin Exp Res, 32*(8), 1388-1397. doi: 10.1111/j.1530-0277.2008.00707.x

Mitchell, M., & Miller, L. S. (2008). Prediction of functional status in older adults: the ecological validity of four Delis-Kaplan Executive Function System tests. *J Clin Exp Neuropsychol, 30*(6), 683-690. doi: 10.1080/13803390701679893

Muckle, G., Jacobson, S., Jacobson, J., LaFlamme, D., Gagnon, J., & Dery, S. (2005). *Substance use during pregnancy among Inuit women from Nunavik [abstract]*. Victoria, B.C.: Canada Northwest Partnership on Fetal Alcohol Spectrum Disorder meeting.

Nash, K., Sheard, E., Rovet, J., & Koren, G. (2008). Understanding fetal alcohol spectrum disorders (FASDs): toward identification of a behavioral phenotype. *ScientificWorldJournal, 8*, 873-882. doi: 10.1100/tsw.2008.75

O'Hare, E. D., Lu, L. H., Houston, S. M., Bookheimer, S. Y., Mattson, S. N., O'Connor, M. J., & Sowell, E. R. (2009). Altered frontal-parietal functioning during verbal working memory in children and adolescents with heavy prenatal alcohol exposure. *Hum Brain Mapp, 30*(10), 3200-3208. doi: 10.1002/hbm.20741

Olson, H. C., Feldman, J. J., Streissguth, A. P., Sampson, P. D., & Bookstein, F. L. (1998). Neuropsychological deficits in adolescents with fetal alcohol syndrome: Clinical findings. *Alcoholism: Clinical & Experimental Research, 22*(9), 1998-2012. doi: 10.1097/00000374-199812000-00016

Ozonoff, S. (1998). Assessment and remediation of executive dysfunction in autism and asperger syndrome. In E. Schopler & G. B. Mesiboy (Eds.), *Asperger syndrome or high-functioning autism? Current issues in autism* (pp. 263-289). New York, NY: Plenum Press.

Parrish, J., Geary, E., Jones, J., Seth, R., Hermann, B., & Seidenberg, M. (2007). Executive functioning in childhood epilepsy: parent-report and cognitive assessment. *Developmental Medicine & Child Neurology, 49*, 412-416. doi: 10.1111/j.1469-8749.2007.00412.x

Pennington, B. F. (1994). The working memory function of the prefrontal cortices: Implications for developmental and individual differences in cognition. In M. M. Haith, J. B. Benson, R. J. Roberts & B. F. Pennington (Eds.), *The development of future-oriented processes*. Chicago IL: University of Chicago Press.

Pennington, B. F., Bennetto, L., McAleer, O., & Roberts, R. J. Jr. (1996). Executive functions and working memory: Theoretical and measurement issues. In G. R.

Luon & N. A. Krasnegor (Eds.), *Attention, memory, and executive function*. Baltimore, MD: Paul H. Brookes Publishing Co.

Petrides, M., & Milner, B. (1982). Deficits on subject-ordered tasks after frontal- and temporal-lobe lesions in man. *Neuropsychologia, 20*(3), 249-262.

Porteus, S. (1959). *The maze test and clinical psychology*. Oxford, England: Pacific Books.

Rasmussen, C. (2005). Executive Functioning and Working Memory in Fetal Alcohol Spectrum Disorder. *Alcoholism: Clinical & Experimental Research, 29*(8), 1359-1367. doi: 10.1097/01.alc.0000175040.91007.d0

Rasmussen, C., & Bisanz, J. (2009). Executive functioning in children with Fetal Alcohol Spectrum Disorders: profiles and age-related differences. *Child Neuropsychol, 15*(3), 201-215. doi: 10.1080/09297040802385400

Rasmussen, C., Horne, K., & Witol, A. (2006). Neurobehavioral functioning in children with fetal alcohol spectrum disorder. *Child Neuropsychol, 12*(6), 453-468. doi: 10.1080/09297040600646854

Rasmussen, C., McAuley, R., & Andrew, G. (2007). Parental ratings of children with fetal alcohol spectrum disorder on the Behavior Rating Inventory of Executive Functioning (BRIEF). *Journal of FAS International*.

Reitan, R. (1958). Validity of the Trail Making Test as an indicator of organic brain damage. *Perceptual and Motor Skills, 8*, 271-276.

Riley, E. P., Infante, M. A., & Warren, K. R. (2011). Fetal alcohol spectrum disorders: an overview. *Neuropsychol Rev, 21*(2), 73-80. doi: 10.1007/s11065-011-9166-x

Robbins, T. W., James, M., Owen, A. M., Sahakian, B. J., McInnes, L., & Rabbitt, P. (1994). Cambridge Neuropsychological Test Automated Battery (CANTAB): A factor analytic study of a large sample of normal elderly volunteers. *Dementia, 5*, 266-281.

Roberts, G. & Nanson, J. (2000). *Best practices: Fetal alcohol syndrome/Fetal alcohol effects and the effects of other substances during pregnancy.* Ottawa: Canada's Drug Strategy Division, Health Canada.

Robinson, G. C., Conry, J. L., & Conry, R. F. (1987). Clinical profile and prevalence of fetal alcohol syndrome in an isolated community in British Columbia. *Canadian Medical Association Journal, 137*, 203-207.

Rothke, S. E. (1986). The role of set shifting cues on the wisconsin card sorting test and halstead category test. *Dissertation Abstracts International, 47*(2-B).

Salimpoor, V. N., & Desrocher, M. (2006). Increasing the utility of EF assessment of executive function in children. *Developmental Disabilities Bulletin, 34*(1&2), 15-42.

Sattler, J. (2008). *Assessment of children: Cognitive foundations* (5th ed.). La Mesa, CA: Jerome M. Sattler.

Sampson, P. D., Streissguth, A. P., Bookstein, F. L., & Barr, H. M. (2000). On categorizations in analyses of alcohol teratogenesis. *Environmental Health Prospectives, 108*(3), 421-428.

Sbordone, R. J. (2000). Ecological validity issues in neuropsychological testing. *Brain Injury Source, 4*, 10-12.

Sbordone, R. J. (2010). Neuropsychological Tests are Poor at Assessing the Frontal Lobes, Executive Functions, and Neurobehavioral Symptoms of Traumatically Brain-Injured Patients. *Psychological Injury and Law, 3*(1), 25-35. doi: 10.1007/s12207-010-9068-x

Sbordone, R. J., & Guilmette, T. (1999). Ecological validity: Prediction of everyday and vocational functioning from neuropsychological test data. In J. J. Sweet (Ed.), *Forensic neuropsychology: Fundamentals in practice* (pp. 227-254). Lisse, Netherlands: Swets & Zeitlinger.

Schonfeld, A. M., Mattson, S. N., Lang, A. R., Delis, D. C., & Riley, E. P. (2001). Verbal and nonverbal fluency in children with heavy prenatal alcohol exposure. *Journal of Studies on Alcohol and Drugs, 62*, 239-246.

Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society of London, Biology, 298*, 199-209.

Shallice, T. (1990). *From neuropsychology to mental structure*. New York: Oxford University Press.

Shear, B. R., & Zumbo, B. D. (2013). False positives in multiple regression: Unanticipated consequences of measurement error in the predictor variables. *Educational and Psychological Measurement, 73*(5), 733-756.

Siklos, S., & Kerns, K. A. (2004). Assessing multitasking in children with ADHD using a modified Six Elements Test. *Archives of Clinical Neuropsychology*, *19*, 347-361.

Square, D. (1997). Fetal alcohol syndrome epidemic on Manitoba reserve. *Canadian Medical Association Journal, 157*, 59-60.

Stevens, J. P. (2009). *Applied Multivariate Statistics for the Social Sciences* (5th ed.). New York: Routledge.

Stratton, K., Howe, C., & Battaglia, F. (1996). *Fetal alcohol syndrome: Diagnosis, epidemiology, prevention, and treatment*. Washington, DC: National Academy.

Stroop, J.R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology: General, 121*, 15-23.

Stuss, D. T., & Alexander, M. P. (2000). Executive functions and the frontal lobes: A conceptual view. *Psychological Research, 63*, 289-298. doi: 10.1007/s004269900007

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

Stuss, D. T., & Knight, R. T. (2002). *Principles of frontal lobe function*. New York: Oxford University Press. Ottawa, Ontario: Aboriginal Healing Foundation.

Tait, C. L. (2003). Fetal alcohol syndrome among Aboriginal people in Canada: Review and analysis of the intergenerational links to residential schools.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). Boston, MA: Allyn & Bacon.

Teunisse, J. P., Roelofs, R. L., Verhoeven, E. W., Cuppen, L., Mol, J., & Berger, H. J. (2012). Flexibility in children with autism spectrum disorders (ASD): inconsistency between neuropsychological tests and parent-based rating scales. *J Clin Exp Neuropsychol, 34*(7), 714-723. doi: 10.1080/13803395.2012.670209

Tien, A. Y., Spevack, T. V., Jones, D. W., Pearlson, G. D., Schlaepfer, T. E., & Strauss, M. E. (1996). Computerized wisconsin card sorting test: comparison with manual administration. *Kaohsiung Journal of Medical Sciences, 12,* 479-485.

Toplak, M. E., Bucciarelli, S. M., Jain, U., & Tannock, R. (2009). Executive functions: performance-based measures and the behavior rating inventory of executive function (BRIEF) in adolescents with attention deficit/hyperactivity disorder (ADHD). *Child Neuropsychol*, *15*(1), 53-72. doi: 10.1080/09297040802070929

Turner, M. (1997). Towards an executive dysfunction account of repetitive behaviour in autism. In J. Russell (Ed.), *Autism as an executive disorder* (pp. 57-100). New York, NY, US: Oxford University Press.

van der Sluis, S., de Jong, P. F., & van der Leij, A. (2004). Inhibition and shifting in children with learning deficits in arithmetic and reading. *J Exp Child Psychol, 87*(3), 239-266. doi: 10.1016/j.jecp.2003.12.002

Vaurio, L., Riley, E. P., & Mattson, S. N. (2008). Differences in executive functioning in children with heavy prenatal alcohol exposure or attention-deficit/hyperactivity disorder. *Journal of the International Neuropsychological Society, 14,* 119-129.

Vik, P., & Ruff, R. R. (1988). Children's figural fluency performance: Development of strategy use. *Developmental Neuropsychology, 4*(1), 63-74.

Vriezen, E. R., & Pigott, S. E. (2002). The relationship between parental report on the BRIEF and performance-based measures of executive function in children with moderate to severe traumatic brain injury. *Child Neuropsychology*, *8*(4), 296-303. doi: 10.1076/chin.8.4.296.13505

Wechsler, D. (1939). *The measurement of adult intelligence*. Oxford, England: Williams & Wilkins.

Wechsler, D. (1949). *Wechsler intelligence scale for children; manual*. Oxford, England: The Psychological Corp.

Wechsler, D. (1991). *The Wechsler Intelligence Scale for Children-Third edition*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (2004). *Wechsler Intelligence Scale for Children-Fourth Edition*. London: Pearson Assessment.

Welsh, M.C., & Huizinga, M. (2001). The development and preliminary validation of the tower of hanoi-revised. *Assessment, 8*(2), 167-176.

Welsh, M.C., & Pennington, B.F. (1988). Assessing frontal lobe functioning in children: Views from developmental psychology. *Developmental Neuropsychology, 4*(3), 199-230. doi: 10.1080/87565648809540405

Yeudall, L. T., Fromm, D., Reddon, J. R., & Stefanyk, W. O. (1986). Normative data stratified by age and sex for 12 neuropsychological tests. *Journal of Clinical Psychology, 42*, 918-946.

VITA AUCTORIS

NAME:	Jaspreet Kaur Rai
PLACE OF BIRTH:	Edmonton, Alberta
YEAR OF BIRTH:	1989
EDUCATION:	University of Alberta, Edmonton, Alberta 2011, B.Sc. (Specialization) Psychology
	University of Windsor, Windsor, Ontario 2014, M.A. Clinical Neuropsychology