2011

An Analysis of the Woodcock-Johnson III and the IVA

Sean Locke
University of Windsor

An Analysis of the Woodcock-Johnson III and the IVA

by

Sean Robert Locke

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

2011

An Analysis of the Woodcock-Johnson III and the IVA

By

Sean Robert Locke

APPROVED BY:

The Department of Psychology

____
Dr. Danielle Soulliere (Outside Reader)

____
Dr. Carlin Miller (Department Reader)

____
Dr. Ken Cramer, Advisor

____
Dr. Jill Jackson, (Chair)

Author's 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

Attention Deficit Hyperactivity Disorder (ADHD) can be assessed using variety of measures including clinical interviews, behavioural observations, respondent checklists, cognitive ability scores, and/or computerized performance tests. The use of cognitive ability tests to help diagnose ADHD has recently gained popularity (Frazier, et al., 2004); the Woodcock-Johnson III Tests of Cognitive Abilities and Tests of Achievement (WJ III) is one such test. However, research on the relationship between the WJ III and ADHD is speculative. The present investigation sought to determine whether the WJ III has an attention factor by factor analyzing WJ III scores and an ADHD measure, the IVA+Plus®. The software application could not converge on a proper solution for the IVA data, thus, the analyses were modified to correlation and regression. Mild support was found for past research on the relationship between the WJ III and ADHD. Although the statistical results of the current investigation were unremarkable, several key areas of research in need of further study were illuminated−the most important of which detected a potential major flaw in a popular commercially-sold ADHD measure.

Dedication

Suppose there is a God. Whether or not you believe, let's make the supposition that there is a perfect and perfectly holy God. Now, suppose He creates a world–not for His own amusement but out of sheer love and inhabits it with animals and plants. Then, for his His crowning act of creation, He made a creature that He shaped in His own image and gave this creature pre-eminence and dominion over the earth.

Now suppose He gave this creature one restriction–just one–not to eat the fruit from a particular tree. And one day, man, who owes their creator everything, grasped for the fruit, grasping for equality with his creator. He gave them one law, for which the penalty was death.

Suppose right then that God had wiped humankind right off of the earth…He would have been perfectly justified to do so, wouldn't He?

But He didn't…Filled with love and compassion He gave them a way. They rejected Him. They were undeserving. Nevertheless, He decides to send His only begotten Son to take on the sins of His people.

And they kill Him–they kill the Son. But God said that's okay, and brought Him back to life, thereby providing a conduit between perfection and imperfection–a way for eternal life.

And God said, believe in Me and put your trust in My Son, and I will forgive you of every sin that you have ever committed and will ever commit against Me.

Do this and I will give you eternal life. The only requirement is that you honour the One who died in your place.

Afterword

If you don't believe, I ask you why? Have you opened a bible? Have you reviewed the thousands upon thousands of sources documents that support the validity and legitimacy of the bible; the same bible that historians consider to be "the most well-supported historical document in existence."

In our line of work, in psychology, untested hypotheses have no validity until tested. When we publish an article, we don't just publish an introduction and hypothesis–we test the hypotheses, with empirical rigor.

Suppose there is a chance that God exists and that Jesus did die for you. Have you done any research, have you done anything to support or reject your view? Or do you simply have a strongly held believe, that has gone unchallenged? An untested hypothesis.

How have you tested your hypotheses about God?

Acknowledgements

St. Andrew's Congregation

Family

Friends

Ken Cramer

Danielle Soulliere

Carlin Miller

God

Greg Chung-Yan

Fuschia Sirois

Barat Wolfe

Kevin McGrew

Corey Yednoroz

Jill Schroeder

Joan Helmsing

Table of Contents

An Analysis of the Woodcock-Johnson III and the IVA

Introduction

The Woodcock-Johnson III Normative Update (WJ III) Tests of Cognitive abilities and Tests of Achievement and its predecessors were developed for psychoeducational assessment. Since its first edition in 1977, the Woodcock-Johnson family of tests have been used by clinicians to help diagnose dyscalculia (inability to do math), dysgraphia (inability to write), dyslexia (inability to read), and auditory, visual, and processing problems, to name a few. With the recent increase in Attention Deficit Hyperactivity Disorder (ADHD) diagnoses (Mandell, Thompson, Weintraub, DeStefano, & Blank, 2005), the WJ III is now being used to supplement other measures in diagnosing ADHD. One such measure is the Integrated Auditory and Visual Continuous Performance Test (IVA+Plus), which follows the DSM-IV diagnostic criteria to help clinicians assess for ADHD.

The present investigation sought to determine whether the WJ III has an attention factor by factor analyzing the WJ III's subtests and latent variables with the IVA's diagnostic quotient scores. This can offer educators and psychologists potential markers of ADHD when administering a WJ III assessment. The proceeding reviews the WJ III and the IVA, then past research on the relationship between cognitive ability and ADHD discussed before the research questions and corresponding hypotheses are purported.

Overview of the WJ III Normative Update and the Cattell-Horn-Carroll Theory of Cognitive Abilities

The Woodcock-Johnson Psychoeducational Battery (WJPEB; Woodcock & Johnson, 1977), predecessor to the WJ III Tests of Cognitive Abilities (Woodcock, McGrew, & Mather 2001), was atheoretical. Instead of a model of cognitive ability, it measured verbal and nonverbal intelligence. The WJ III (Woodcock, McGrew, & Mather, 2001) is the most recent edition of the Woodcock-Johnson batteries. It is based on the Cattell-Horn-Carroll (CHC) theory of cognitive abilities (Schrank et al., 2002). The WJ III was published in 2001 and had a normative update (NU) in 2007. The original 2001 WJ III norms were based on year 2000 U.S. Census projections available at the time the standardization of the WJ III commenced in 1996. Census projections are estimates of the population for future dates and are subsequently replaced by census statistics. The final 2000 census statistics produced a somewhat different description of the U.S. population than was available from the last projections issued in 1996. The WJ III NU updated the WJ III norms to reflect the final U.S. 2000 census statistics. In addition, bootstrap re-sampling methods were used in the development of the WJ III NU norms, these methods were not fully developed at the time of the 2001 publication of the WJ III.

McGrew (1997) was the first to synthesize Cattell-Horn's *Gf-Gc* theory of fluid and crystalized intelligence and Carroll's Three-Stratum models in an attempt to provide a comprehensive and integrative framework for interpreting human cognitive abilities. The result is the CHC theory, which served as the theoretical blueprint for the WJ III (McGrew & Woodcock, 2001). An understanding of the CHC model is needed, in order to appropriately interpret the WJ III.

The CHC model incorporates Carroll's (1993) Three-Stratum theory of intelligence, organizing cognitive abilities and Cattell -Horn Gf-Gc theory into an integrated three level hierarchy. Carroll (1993, 1997) identified over 69 specific, or narrow cognitive abilities, at Stratum I. The narrow abilities are subsumed under the broad (Stratum II) cognitive ability domains of Fluid Intelligence or Reasoning (Gf), Crystallized Intelligence or Comprehension-Knowledge (Gc), Broad Visual-Spatial Processing (Gv), Broad Auditory Processing (Gu), and

Processing Speed (Gs). At the apex of his model (Stratum III), Carroll identified a higher-order factor above the broad factors, which he interpreted as General Intelligence, or *g* (for an extensive discussion of the CHC model and Carroll's Three-Stratum theory see Carroll (1993) and McGrew (2005, 2009).

Although CHC theory synergistically bridges *Gf*-*Gc* theory and Three-Stratum theory, there are inconsistencies between these two etiological theories. McGrew (1997) outlines the primary similarities and differences between Cattell-Horn's and Carroll's theories of intelligence. Both theories concur that there are several narrow cognitive abilities, but they do not agree on the exact number of abilities (Carroll − 69 abilities; Cattell-Horn − 87 abilities). Also, both models incorporate the broad abilities of *Gf, Gc, Ga, Gv*, and *Gs*. However, while Cattell-Horn separates *Gsm* and *Glr*, Carroll combines them (in the General Learning and Memory (*Gy*) and Broad Retrieval (*Gr*) abilities). The main discrepancy between the two parent theories is the presence (Carroll) or absence (Cattell-Horn) of a general intelligence (*g*) factor at Stratum III. This last point has provoked debate as an irreconcilable inconsistency between Carroll's and Cattell-Horn's view of intelligence.

To counter critics, McGrew (2005) emphasizes that the empirical evidence heavily weighs in favor of the presence of a *g* factor. Although biased towards the presence of *g,* Carroll (2003) concluded that "researchers who are concerned with this structure [of CHC theory] in one way or another … can be assured that a general factor *g* exists" (p. 19).Thus, the CHC theory is not just based on a single notion of intelligence (g), rather on a blend of the theories of Cattell, Horn, and Carroll. This theory recognizes that intelligence cannot be viewed simply as a single factor, but that intelligence is much more complex and represents a multifaceted notion of intelligence (McGrew & Woodcock, 2001). The WJ III uses CHC theory by providing a measure of fluid reasoning (*Gf*), comprehension knowledge (*Gc*), short-term memory (*Gsm*), long-term retrieval (*Glr*), processing speed (*Gs*), visual-spatial thinking (*Gv*), and auditory processing (*Ga*). The WJ III COG and ACH provide for additional combinations of tests that are useful in diagnosing or identifying an individual's strengths and weaknesses and include areas of Phonemic Awareness, Working Memory, Cognitive Fluency, and Executive Processes (Mather & Woodcock, 2001).

The WJ III NU (Woodcock, McGrew, Schrank, & Mather, 2001; 2007) is a comprehensive measure of cognitive abilities and achievement organized into three distinct, co-normed test batteries: The Woodcock-Johnson III NU Tests of Cognitive Abilities (WJ III COG); the Woodcock-Johnson III Diagnostic Supplement to the Tests of Cognitive Abilities (WJ III DS) and the Woodcock-Johnson III NU Tests of Achievement (WJ III ACH). The WJ III is designed to measure a wide array of cognitive, oral language, and academic achievement abilities for individuals from preschool (2 years) through the geriatric (90+ years) age levels.

Both the WJ III COG and ACH are organized into Standard and Extended batteries that can be used independently, together, or in conjunction with other tests (including tests from the WJ III DS). In addition to the CHC clusters, the complete set of WJ III NU thirty-one (20 in original WJ III Cognitive battery plus 11 additional in the WJ III DS) cognitive tests are also organized by three broader categories related to cognitive performance (Cognitive Performance Model, Woodcock, 1997): Verbal Ability, Thinking Ability, and Cognitive Efficiency and five clinical clusters (Broad Attention, Executive Functioning, Working Memory, Cognitive Efficiency, and Phonemic Awareness). Twenty-two achievement tests are organized by curricular area (reading, mathematics, written language, and academic knowledge) and oral language and by clusters within these areas (e.g. Basic Reading Skills, Math Reasoning), with additional groupings for special purpose clusters (e.g. Academic Skills, Phoneme/Grapheme Knowledge). These batteries have particular diagnostic utility in that examiners are encouraged to be selective in their testing and select different evaluation tools based on the unique circumstances of each referral. Appendix A and Appendix B contain a more thorough description of the WJ III COG and ACH tests and their corresponding CHC cluster that were used for the current investigation.

Like the earlier versions of the WJ, the WJ III has been viewed as state of the art in the individual measurement of cognitive abilities and achievement (Cizek, 2003; Cummings, 1995; Hicks & Bolan, 1996; Lee & Stefany, 1995; Sandoval, 2003). The WJ has long been one of the most widely used individually administered academic achievement batteries (Kamphaus, Windsor, Rowe, & Kim, 2005). Furthermore, the WJ III COG is taught as a measure of intelligence in many school psychology post-graduate programs across the U.S. and Canada (Braden & Alfonzo, 2003; Ford, Percy, & Negreiros, 2010). Its strong psychometric properties, the co-normed tests of cognitive abilities and achievement, its utility for use with individuals throughout the lifespan, and features which assist in understanding unique processing strengths and weakness, contribute to its increasing use in Canada.

The WJ III in Canada

The WJ III is used throughout Canada, however, the question arises as to whether a U.S. normed test is appropriate for use in Canadian populations. Whereas a number of U.S. normed batteries of cognitive and achievement abilities are used extensively throughout Canada, surprisingly few comprehensive validation and/or standardization studies are reported in the literature. Of the limited published studies that address the use of cognitive and achievement with U.S. norm used in Canada, the majority published to date have examined differences in the various versions and editions of the Wechsler scales standardized in the U.S. and administered to Canadian populations. All have pointed to significant score differences across the Canadian and U.S. populations with Canadians samples scoring on average 2 to 5 standard score points higher than the U.S. sample depending on the factor, or subtest (Hildebrand, & Saklofske, 1996; Wechsler, 1996; 2001; 2003; 2004; 2008). These findings have suggested the need for Canadian standardization of the Wechsler scales.

Given the widespread use of many US-normed cognitive and achievement batteries with Canadian populations for diagnosis, treatment, and program planning, research is needed to determine whether the U.S. norms are "transportable" and applicable to Canadian populations. A review of the literature reveals that, to date, only two WJ III studies have evaluated the transportability of the WJ III NU norms across the border for use with school-aged Canadian students (Ford, Swat, Negreiros, Lacroix, & McGrew, 2010; Locke, McGrew, & Ford, 2011).

In their study, Ford and Swart et al. (2010) found that although the U.S. sample typically scored slightly higher than the Canadian sample on the WJ III NU COG clusters, the differences were not statistically significant, with one exception the Long-Term Retrieval cluster. The mean difference of the General Intellectual Ability-Extended (GIA-Ext) score for the U.S. sample (M = 100.74, SD = 15.77) was not significant (although it was near significance; t (309) = 1.84, p > .05) compared to for the Canadian sample (M = 98.88, SD = 13.73). While the Canadian sample scored slightly higher (M = 101.30, SD = 14.16) than the U.S. sample (M = 100.90, SD = 15.37) on the Total Achievement cluster the difference was not a statistically significant, t *(309)* = .37, *p=.715.* However, at the test level, five statistically significant differences were noted: the Canadian sample scored significantly higher on the Reading Fluency, Quantitative Concepts, and Oral Comprehension tests while the U.S. sample scored statistically higher on the Reading

Vocabulary and Editing tests. Thus, these findings support, with some caution, the use or transportability of the *WJ III NU* U.S. based norms with Canadian populations.

These results lay the foundation for further research on the portability of the U.S. normed WJ III to Canadian populations. T-tests reveal whether there is a large enough difference between the means of two samples to be considered a statistically significant difference. However, t-test cannot interpret the meaning of scores or determine whether the scores mean the same thing; for that we turn to factorial construct validity tests, such as a confirmatory factor analysis (CFA).

In support of the construct validity of the WJ III CHC measurement model, McGrew and Woodcock (2001) presented an extensive set of CFA across five broad age groups (spanning ages 6 through 100), as well as for a combined sample across all ages. Taub and McGrew (2004) also reported support for the invariance of the CHC factor structure of the WJ III throughout the entire age range (ages 6 to 100 years). However, this research was conducted in the United States; thus, Locke, McGrew, and Ford (Submitted) examined the comparability of the WJ III NU's factor structure in matched school-age Canadian/U.S. samples. The study investigated the measurement invariance of the WJ III between matched Canadian/U.S. samples; the results supported the metric invariance of the WJ III across the two countries.

With the validation of the WJ III's U.S. norms and factor structure in Canada, the present study uses a sample of students and adults from Western Canada to explore whether there are WJ III subtests or clusters that can predict ADHD.

Attention Deficit Hyperactivity Disorder (ADHD)

Attention deficit hyperactivity disorder (ADHD) is a common neurodevelopmental disorder of childhood that may be associated with substantial cognitive, social, behavioural, and academic impairment. ADHD affects 3-7% of the school-aged population (American Psychiatric Association, 2000; Barkley R. A., 1998; Hoff, Doepke, & Landau, 2002). The National Comorbidity Survey Replication Study found 36% of adults who met the criteria for ADHD as children continued to meet the diagnostic criteria as adults (Kessler, Adler, Barkley, & et al., 2005). The overall prevalence of adults who meet the diagnosis of ADHD is 4.4% (Kessler R. C., Adler, Barkley, & et al., 2006). To assist in the understanding of ADHD, Appendix C describes the Diagnostic and Statistical Manual of Mental Disorders, fourth edition (DSM-IV) diagnostic criteria for ADHD and its three associated subtypes (American Psychiatric Association, 2000).

Clinical assessment of ADHD varies greatly from one clinician to another. Some are inclined to evaluate a client based on one source of information (e.g. an unstructured interview), while others might use a comprehensive assessment, including cognitive and neuropsychological testing, achievement testing, behavioural observations, and structured diagnostic interviews. Obviously comprehensive assessments are generally preferred to single-informant reports because the additional information gathered helps to rule out the diagnoses of alternative explanations for the pattern of symptoms (Barkley, 1998). Although neuropsychological instruments are often used in ADHD testing, there is little agreement regarding their utility (Rapport, Chung, Shore, Denney, & Isaacs, 2000).

The American Academy of Child and Adolescent Psychiatry (AACAP) established practice parameters for the diagnosis and treatment of ADHD (Pliszka and AACAP Work Group on Quality Issues, 2007). They agree that the best practice for diagnosing ADHD focuses on gathering information from multiple sources, as previously described. These standards have high degree of similarity to Kamphaus and Campbell’s (2006) model of diagnosing ADHD, lending further support for a multi-etiological approach; its reference in several psychological and educational resources for ADHD (e.g. Kamphaus and Campbell, 2006; Miller, 2010) also supports AACAP's model. However, AACAP discourages the use of computerized CPTs as they "are not generally useful in diagnosis" (Pliszka and AACAP Work Group on Quality Issues, 2007). A further discussion of the diagnostic utility of CPTs (i.e. the IVA) can be found in the limitations and future opportunities section.

In such comprehensive assessments, a neuropsychological or cognitive assessment is a typical starting point to evaluate the individual's cognitive ability (Frazier, Demaree, & Youngstrom, 2004). However, research regarding the use of intelligence tests as a means of distinguishing between individuals with and without ADHD is divided. While some studies find up to 20-point discrepancies between ADHD and non-ADHD groups (Abikoff, Courtney, Szaibel, & Koplewicz, 1996; Garcia-Sanchez, Estevez Gonzalez, Suarez Romero, & Junque, 1997), others have found little to no difference (Carelson & Tamm, 2000; Carleson, Mann, & Alexander, 2000). Thus, there is utility in establishing whether a cognitive or neuropsychological battery can predict ADHD.

Carroll (1993) clearly articulated the unknown relationship in a model of cognitive abilities when he stated:

> "...it can be argued that attention is involved, in varying degrees, in all cognitive performances and, thus, in all performances that are regarded as indicating cognitive abilities. One can expect it to be very difficult to separate the attentional components of such performances from those components that represent latent traits of abilities other than the ability to attend. An individual differences factor could often be equally well

interpreted either as a factor of some particular cognitive ability or as a factor of attentional ability. (p. 547)"

To address this concern, a recent study conducted an exploratory structural equation model on 17 attentional tests and 14 tests of cognitive ability designed to measure constructs in Carroll's taxonomy of intelligence (Burns, Nettelbeck, & McPherson, 2009). Three separate attention factors were identified: two were interpreted as reflecting aspects of sustained attention (one being conceptually similar to *Gs* abilities) and one working-memory capacity factor (*Gsm*). Practically this suggests that many speeded cognitive tests on intelligence batteries may be reflecting the strong influence of sustained attention (as suggested in Carroll's quote). The other sustained attention factor might be getting at a more "attentional" construct as it demanded that performance be maintained for longer periods or with more complex tests, or both as opposed to continuous speeded performance (Burns et al., 2009).

Cognitive Batteries and Predicting ADHD

This section is prefaced with a few opening remarks. Nearly all of the research that has been conducted on the diagnostic utility of cognitive assessments used for predicting ADHD has focused on the Wechsler family of tests in their analyses. Thus, the reviewed literature primarily discusses the Wechsler family of tests; however, comparative studies have been conducted to analyze the similarities between subtests of the WJ III and the Wechsler family of tests. For the purpose of this investigation, this literature helps facilitate our understanding of which WJ III subtests are related to ADHD.

Although the diagnostic criteria for ADHD are primarily behavioural manifestations, cognitive deficits such as attentional impairments, response inhibition, and perceptual-motor speed have been hypothesized to be the central features of the disorder (Barkley, DuPaul, &

McMurray, 1990). Despite the fact that there is a dispute over the etiology of ADHD, there is empirical support demonstrating that ADHD has cognitive and a behavioural components (Kronenberger & Meyer, 1996). For example, Barkley, Grodzinsky, and DuPaul (1992) argue that some children with ADHD have cognitive deficits in attention and inhibitory control; Lahey et al. (1988) found that the perceptual motor speed of children with ADHD was impaired. Barkley (1997) proposed that ADHD reflects response inhibition, which is related to impaired executive functioning in four specific subareas: self-regulation of affect, internalized speech, working memory, and reconstitution or verbal fluency. One meta-analysis of ADHD assessments found that scores on sustained attention (Digit Span, Coding, Symbol Search), working memory (Digit Span and Arithmetic), and verbal fluency (Letter Fluency and Category Fluency) were significantly smaller in those with ADHD than those without (Frazier, Demaree, & Youngstrom, 2004).

A variety of approaches have been used to identify the utility of the intelligence tests (namely the Wechsler Intelligence Scale for Children (WISC)) in the identification of children with ADHD. One approach has involved comparing the mean cognitive performance scores of the Wechsler scales in children with and without ADHD (Schwean & Saklofske, 1998). As previously mentioned, research is divided as to how the two groups compare; it seems as though more literature points to the non-ADHD children performing better than their ADHD counterparts (e.g. meta-analyses by Frazier et al., 2004 & Bridgett & Walker, 2006). Mean-level comparisons of cognitive ability between individuals with and without ADHD have provided valuable descriptive information about children with ADHD; however, it has been argued that mean-level comparisons on neuropsychological tests have little diagnostic utility as performance patterns of children with ADHD can be similar to patterns exhibited by children and adolescents with other organic, mental, and educational disorders (Frazier, Youngstrom, Glutting, & Watkins, 2007).

Profile analyses have also been used to gain information about the diagnostic utility of the WISC-III with ADHD populations (Prifitera & Dersh, 1993). It has been found that children with ADHD reliably score lower on the Arithmetic, Coding, Information, and Digit Span (ACID) subtests (Prifitera & Dersh, 1993). Thus, a low ACID profile would indicate the potential for ADHD. The problem with the ACID profile and other types of profiles are that they are more effective at ruling out a diagnosis of ADHD than predicting it. Once again mean scores are used to differentiate between groups, but do not address questions of prediction and classification.

The ACID profile was created and based-off of the heavily criticized DSM-III-R's diagnostic criteria for ADHD because much of its research is not valid under the contemporary DSM-IV diagnostic criteria (America Psychological Association, 2000). The primary change in the DSM-IV was the re-establishment of the three core diagnostic dimensions: inattention, impulsivity, and hyperactivity, which the DSM-III-R eliminated (Kamphaus & Frick, 2005). Also, the DSM-IV eliminated the DSM-III-R's diagnostic category of Attention Deficit Disorder with Hyperactivity (ADD/H) and without Hyperactivity (ADD-WO) as they ended up being poor diagnostic categorizations. Thus, caution should be exercised when discussing the utility of the ACID profile.

A third approach, discriminant function analyses (DFA), has been used to differentially diagnose groups of children with and without ADHD using Wechsler scores (Assesmany, McIntosh, Phelps, & Rizza, 2001; Wielkiewicz, 1990). DFA examines the underlying dimensions that differentiate two groups−in this case ADHD from non-ADHD. However, results have generally found that the WISC-Revised's Freedom from Distractibility factor did not significantly discriminate between individuals with and without ADHD (Krane & Tannock, 2001). Rather than diagnosing ADHD, the Freedom from Distractibility factor signals the presence of a learning problem. Using DFA, Assesmany et al. (2001) found that Digit Span, Information, Vocabulary, and Picture Completion on the WISC-III significantly contributed to the prediction of ADHD. With the use of these tests 90% of children with ADHD were classified correctly as well as 87.5% of non-ADHD children were correctly identified.

The Integrated Visual and Auditory Continuous Performance Test

The Intermediate Visual and Auditory is a 13-minute Continual Performance Test (CPT). It was created by Dr. Joseph Sanford and Dr. Ann Turner and is distributed by BrainTrain (BrainTrain, 2010). The IVA+Plus (henceforth referred to as IVA) is the most recent version of the program. The subject must click the mouse only when he or she sees or hears the number “1” and refrains from responding when he or she hears or sees the number “2”. Using normative data from 781 subjects (423 female, 358 male), at 10 different age groupings, the test is designed to assess two major factors: response control and Attention. In addition, the IVA provides an objective measure of fine motor hyperactivity. The manual states that the program is useful for persons between the ages of 5 and 90+ (Sandford & Turner, 1995). Among the many variables are six core quotients and 22 subscales. One issue associated with the IVA is its poor test-retest reliability (.37 to .75) for its composite variables. The IVA contains three validity scales, which ensure scores in ADHD ranges come from ADHD behaviours and not motor problems, fatigue, or random answering (Sandford & Turner, 1995; 1994).

The IVA’s test scores are set at a mean of 100 and a standard deviation of 15. The IVA is divided into four diagnostic categories: attention, response control, attribute, and symptomatic.

It has two primary diagnostic scales: the full scale response control quotient and the full scale attention quotient.

The full scale response control quotient is based on individual auditory and visual response control quotient scores. They are derived from visual and auditory prudence, consistency and stamina scales. Prudence is a measure of impulsivity and response inhibition as seen by three different types of errors of commission. Consistency measures the general ability and variability of response times and is used to help measure one's ability to stay on task. Stamina compares the mean reaction times of correct responses during the first 200 trials and the last 200 trials. These scores are used to identify problems related to sustaining attention and effort over a period of time.

The full scale attention quotient is derived from separate auditory and visual attention quotients. The attention quotient scores are based on equal measures of visual and auditory vigilance, focus, and speed. Vigilance is a measure of inattention as evidenced by two different types of errors of omission. Focus reflects the total variability of mental processing speed for all correct responses. Speed reflects the average reaction time for all correct responses throughout the test and helps to identify attention processing problems related to slow discriminatory mental processing.

The fine motor regulation scale provides additional information by recording off-task behaviours with the mouse such as: multiple clicks, spontaneous clicks during instruction periods, anticipatory clicks, and holding the mouse button down. In behavioural terms, fine motor regulation scores quantify fidgetiness and restlessness associated with small motor hyperactivity. The balance and readiness scores provide clinicians with data regarding the client's learning style. Balance indicates whether the test taker processes information more quickly visually or aurally, or is uniformly consistent in both modalities. Readiness indicates whether the test taker processes information more rapidly when the demand is quicker or when it is slower. It provides a subtle measure of inattention when the test taker just “can’t quite keep up” (BrainTrain, 2010).

IVA+Plus’ Symptomatic scales include auditory and visual comprehension, persistence, and sensory/motor. Comprehension identifies random responding by measuring idiopathic errors. Research has shown this to be the single most sensitive sub-scale in discriminating ADHD. Persistence is a measure of motivation when the test taker is asked to do one more thing at once. It can also reflect motor or mental fatigue. Sensory/motor scales provide a measure of reaction time speed to simple, singular test stimuli (i.e. the “2”). These scales help screen for slow reaction times which may impair test performance or possibly indicate neurological, psychological or learning problems not associated with ADHD. Although the IVA+Plus scores are divided into four categories, attention, response control, attribute, and validity, the attention and response categories are the primary diagnostic scales (Sandford and Turner, 1995).

The WJ III Subtests and Clusters that may be related to ADHD

A meta-analysis of ADHD assessments and cognitive ability (using the WISC) revealed that sustained attention, working memory, and verbal fluency differentiated ADHD from non-ADHD individuals (Frazier et al., 2004). Sustained attention is measured by Digit Span, Coding, and Symbol search on the WISC-III. The WJ III’s Memory for Words (*Gsm*) is equivalent tests to Digit Span, Visual Matching (*Gs*) is equivalent to Coding, and Pair Cancellation (*Gs*) is equivalent to Symbol Search. Equivalence between a WISC-III and WJ III subtest indicate that the two tests load on the same CHC broad cluster (Phelps, McGrew, Knopik, & Ford, 2005) and that they are operationalized in a similar manner. For example, visual matching (WJ III) and coding (WISC-III) both measure the ability to rapidly discriminate and identify two identical numbers (WJ III) or symbols (WISC-III) on a line of numbers within a 2-minute timespan.

The WISC-IV measures working memory using Digit Span and Arithmetics. The WJ III's Short-Term Memory cluster (*Gsm*) was operationalized to subsume the CHC narrow ability of Working Memory (Flanagan, McGrew, & Ortiz, 2000; McGrew & Woodcock, 2001). Specifically Numbers Reversed and Memory for Words are used to measure Short Term Memory on the WJ III. The WISC-IV measures Verbal Fluency using Letter Fluency and Category Fluency. Equivalent tests of the WJ III include Letter-Word Identification (*Grw*) and Reading Fluency (*Grw*).

Executive functioning deficits may exist in individuals with ADHD and might provide an avenue for distinguishing individuals with and without ADHD. Executive functions are frequently described as the mental operations that promote the organization of thought and behaviour. Barkley (1997) developed a theoretical model that links inhibition to four executive neuropsychological functions that depend on it for their effective execution: working memory, self-regulation of affect, internalized speech, and reconstitution (verbal fluency). Applied to ADHD, the model predicts that ADHD should be associated with secondary impairments in these four functions as well as motor control (Barkley, 1997).

Floyd, Shaver, and McGrew (2003) found a set of cluster scores on the WJ III COG, called *clinical clusters*, above and beyond the proposed CHC clusters. These clinical clusters were designed to provide measures of working memory, cognitive speed, attention, and executive processes. In a later study, Floyd et al. (2006) found statistically significant, positive, moderate relationships between these four clinical clusters and the empirically tested Delis-Kaplan Executive Function System. Thus, these four clinical clusters and their associated subtests might be impaired in those with ADHD; they are related to executive functioning, which is typically found as a deficit in individuals with ADHD.

Floyd et al. (2006) suggest that the Working Memory cluster was developed to measure the phonological and central executive components of the memory management system called working memory. This cluster results from performance on the Numbers Reversed and Auditory Working Memory tests. The Cognitive Fluency cluster was developed to measure the speed at which individuals retrieve information from memory, produce words, and make decisions about conceptual similarities. This cluster results from performance on the Retrieval Fluency, Decision Speed, and Rapid Picture Naming tests.

The Broad Attention and Executive Processes clusters were developed to measure abilities related to executive functions. The Broad Attention cluster was designed to provide a general measure of attention by tapping into qualitatively different aspects of the construct. This cluster results from performance on the Numbers Reversed, Auditory Working Memory, Auditory Attention, and Pair Cancellation tests. The Executive Processes cluster was developed to measure the core cognitive processes associated with executive functions, such as response inhibition, cognitive flexibility, and planning. This cluster results from performance on the Concept Formation, Planning, and Pair Cancellation tests.

There have been three small-scale, non-peer-reviewed studies that examined potential indicators of ADHD using the WJ III. Two of the studies (Vesley, 2001 & Lerner and Yasutake, 2001) were published in the Woodcock-Johnson III Training Manual, while the other (Ford, Keith, Floyd, Fields, and Schrank, 2003) was published in the Woodcock-Johnson III Assessment and Interpretation Manual. Vesley (2001) found that Visual-Auditory Learning, Concept Formation, Visual Matching, Incomplete Words, Auditory Working Memory, Decision Speed, and Rapid

Picture Naming are tests that may be sensitive to ADHD. Lerner and Yasutake (2001) found that Auditory Attention, Calculation, Passage Comprehension, and Oral Comprehension are tests may be sensitive to ADHD. Ford and Keith et al. (2003) found that Concept formation, Numbers Reversed, Auditory Working Memory, Planning, and Pair Cancellation are tests that may be sensitive to ADHD. From these three studies, it was found that the broad cognitive abilities of Short-Term Memory (*gsm*), Processing Speed (*Gs*), and Fluid Intelligence (*Gf*) may be related to measures of attention.

The Present Study

Review of the literature suggests a lack of research that examines the relationship between the WJ III and ADHD, as evidence is speculative; for instance, the discussion of the congruent WJ III/WISC-IV subtests only shows the need for further research. The review of the executive functioning deficits in individuals with ADHD and the review of the WJ III's clinical clusters provide some substantiation that certain tests and clusters of WJ III might be related to measures of attention and of ADHD. The three studies on the WJ III and ADHD have not gone through the empirical rigor of peer review and their conclusions are speculative. Thus, the present investigation seeks to expand our current knowledge of the WJ III's relationship to attention and ADHD beyond inferential statements and toward empirically founded evidence.

Research Question: Are there attention factors on the Woodcock-Johnson III that relate to the IVA's measures of ADHD?

Hypothesis 1 (at Stratum I): The WJ III tests of Memory for Words, Numbers Reversed, Visual Matching, Visual Auditory Learning, Passage Comprehension, Oral Comprehension, Pair Cancellation, Concept Formation, Reading Fluency, Auditory

Attention, General Information, Math Fluency, and Decision Speed will load highly onto the IVA's factors of ADHD.

Hypothesis 2 (at Stratum II): The WJ III clusters of Short-Term Memory (*gsm*), Processing Speed (*Gs*), and Fluid Intelligence (*Gf*)) will load highly onto the IVA's factors of ADHD.

Hypothesis 3 (at Stratum III): The WJ III construct of general intellectual ability (*g)* will not correlate to the IVA's measures of attention and of ADHD.

Do the Woodcock-Johnson III subtests load on the Integrated Visual and Auditory Continuous Performance Test factors using a confirmatory factor analysis? Those WJ III factors that load onto the IVA's factors indicate that they are related to measures of attention or ADHD.

Method

Participants and Data Archive

Participants for the current investigation were taken from a clinical psychologist's archive of educational assessments in Central Canada. Participants include both children and adults. Since the database was archival, the only demographic information available was sex and grade. The final database contained n = 156 participants. There were 70 females and 86 males. The average grade of the sample was M = 10.8 (SD = 4.6); about one-third (n = 50) of the participants were in elementary school, about one-third (n = 51) of the participants were in high school, and the remaining one-third (n = 55) were adults that had graduated high school.

The psychologist's practice standards included gaining consent from the client or client's parent prior to the assessment. Some clients were referred from their grade school for assessment to measure cognitive ability or to assess for ADHD, while others were referred to the psychologist's private practice for various cognitive, attentional, and emotional assessments.

Those clients who required medications (e.g. for ADHD, for psychological problems, or for medical reasons) were assessed on their medications. This practice is important as many−especially those with ADHD−should function optimally while on their medication. However, some clinicians assess individuals off their medication, citing that medications can confound test performance.

Practice standards for assessing ADHD included gathering information about the client as well as from relevant stakeholders such as, parents, teachers, and employers (if applicable). Previous personal, educational, vocational, and medical history was routinely gathered. At the time of testing, depression and anxiety inventories were also used. Other measures were also collected at the time of the assessment that were beyond the scope of the present project.

The assessment database was compiled from two different sources, the psychologist's private practice and six private schools. The data were collected from an archive of over 4000 hard copies of the educational assessments. However, only about 10% of these were used for the current investigation because the remaining 90% used previous editions of the Woodcock-Johnson, which have different norming populations than the WJ III. The hard copies were scanned and saved in .pdf format. Using a double-data-entry system, data were entered into a SPSS database. Using a computer program designed for double data entry, the first assistant entered the information into the database. The second assistant later entered the same information into the same database. When there was an inconsistency between the two assistants input, the data point was flagged so it could be correctly entered. Participants were each assigned a code to ensure the database remained anonymous. The .pdf files and the database were both saved on secure hard drives.

Measures

Non-identifying demographic information was collected. The WJ III and the IVA were administered to each client as part of their educational assessment, along with other varying tests (e.g. depression or anxiety measures) depending on the client's reason for referral. Please refer to the literature review section for an in-depth discussion of the WJ III COG and ACH subtests. Please refer to Appendix A and B for a complete list of the subtests that were used in the current investigation.

The IVA+Plus was also administered to the clients. The 13-minute computer continuous performance test was administered to every client in the current study. For a more in-depth discussion of the IVA+Plus please refer to the aforementioned review of the IVA in the literature review section.

Analyses

A confirmatory factor analysis (CFA) was conducted to determine if a separate attention factor emerged when combining the WJ III and IVA test variables and latent constructs. The Canadian WJ III's factor structure was developed using 30 of the WJ III's subtests (Locke et al., 2011), however, only 25 of the subtests were available in the current investigation, due to a lack of sample size (e.g. Editing had 9 data-points).

As opposed to a research database, clinically-derived databases do not typically administer every subtest to every person. Thus, there were several non-random missing data-points and 94 list-wise deletions were made to get to a sample size of n = 158. T-tests were run to determine whether the removed 94 participants had statistically significant differences from those not removed in demographics or in cognitive ability.

The initial step in the analyses investigated the number of first- and second order factors of the WJ III and of the IVA; in other words, does this particular database fit the specified WJ III and IVA factor structures? While the WJ III was specified to match its proposed factor structure (Locke et al., 2011), two competing models of the IVA were specified in this step: the model proposed by Sandford and Turner (1995) containing full scale response and attention quotients and a second model containing visual and auditory response quotients. The rationale for specifying a competing model was grounded in logic and was a result of the IVA's low internal consistency reliability (Buros Institute, 2007). Prior to specifying the measurement and structure models, it is important to verify that the data fits the two individual factor structures.

The next planned step involved specifying models. In CFA, two types of models are specified: measurement models and structural models. Measurement models are designed to provide a base of comparison by specifying the fit of the model without any specified paths, thus, specifying the fit of the uncorrelated WJ III and the IVA. Structural models represent the fit of the model with the specified *a prior* correlations. Comparing the two allows for the assessment of the X^2 difference, which the degree to which the hypothesized correlations between the two models fits.

The statistical significance of each model was tested via the obtained X^2 goodness-of-fit statistic (cutoff set at determined p value). The results from these analyses were also evaluated using goodness of fit indices that provide empirical evidence of the degree of correspondence between the proposed theoretical model and the data. Since different fit indices are attuned to different aspects of fit, various indices were used in the analyses. To assess the fit of a single model, Hu and Bentler's (1999) "rules of thumb" were used for determining whether retain or reject a model as being plausible. Namely, these conventions are the Tucker-Lewis Index (TLI, also called the non-normed fit index; sensitive to model complexity), and the Comparative Fit Index (CFI; sensitive to non-centrality). Values for the CFI and TLI indices can range from zero to unity, with values of > .95 indicating an excellent fit and fit indices > .90 indicating an adequate fit (Hu & Bentler, 1999). Amos 7.0® (Arbuckle, 2006) was used for these and all CFA analyses.

A final fit index, the Root Mean Square Error of Approximation (RMSEA) statistic takes into account the error of approximation in the population and answers the question “How well would the model, with unknown but optimally chosen parameter values, fit the population covariance matrix if it were available?” (Browne & Cudek, 1989; p. 137-138). Additional advantages of the RMSEA are (a) its sensitivity to the number of estimated model parameters (model complexity) and (b) the provision of 90% confidence intervals that assess the precision of the RMSEA estimates (Byrne, 2001). RMSEA values range from 0.00 to 1.00 with zero indicating no error (a perfect fit). Typically, RMSEA values equal to or less than .05 indicates good fit and values up to .10 suggest adequate or mediocre fit (Byrne, 2001). All fit indices and cut-offs were established *a priori*.

The software application could not converge on a proper solution. This was evidenced by paths (standardized regression weights) exceeding one, by only three significant paths in the model out of the possible ten, by negative error variances, and by several near zero correlations. Thus, the data were analyzed using a different type of analyses. Stepwise regressions were conducted to determine the relationship between the WJ III’s subtests and latent variables and the IVA’s two diagnostic quotient scores: the full scale attention quotient and the full scale response quotient. Since uncorrelated variables cannot reliably predict one another, bivariate correlations were run prior to running the regressions. As a result of the analysis modification, the language and focus of the hypotheses warranted adjustment.

Re-specified Hypotheses

Hypothesis 1 (at Stratum I): The WJ III tests of Memory for Words, Numbers Reversed, Visual Matching, Visual Auditory Learning, Passage Comprehension, Oral Comprehension, Pair Cancellation, Concept Formation, Reading Fluency, Auditory Attention, General Information, Math Fluency, and Decision Speed will correlate and predict the IVA's two full scale quotient factors.

Hypothesis 2 (at Stratum II): The WJ III clusters of Short-Term Memory (*gsm*), Processing Speed (*Gs*), and Fluid Intelligence (*Gf*) will correlate and predict the IVA's two full scale quotient factors.

Hypothesis 3 (at Stratum III): The WJ III construct of general intellectual ability (*g)* will not correlate and predict the IVA's two full scale quotient factors.

Results

As aforementioned, 94 listwise deletions were made prior to the analyses. The list-wise deletions appeared to comprise a subset of cases that were all not administered the same five WJ III subtests. This means that the data was not missing at random. *T*-tests were used to assess whether the 94 cases that were removed significantly differed from the cases that were kept, indicating whether or not the sample was biased. The proportion of males and females did not significantly differ, $t(250) = .387$, $p > .05$) between the removed cases ($M = .460$) and the kept cases ($M = .466$). Average general intellectual ability did not significantly differ, $t(248) = .201$, $p > .05$ between the cases removed ($M = 91.5$, SD = 12.21) and those kept ($M = 93.45$, SD =

10.10). However, there was a statistically significant difference ($t(250) = 3.623$, $p < .001$) in average grade between the cases removed ($M = 13.0$, SD = 3.8) and those cases kept ($M = 10.8$, SD = 4.6). T-tests were not run on the WJ III subtests or the first-order cognitive abilities, as the number of t-tests would have greatly diminished power.

Assumptions were tested to assess how tenable the results of the current analyses are. The violation of an assumption can affect the power of the analyses and the generalizability of the conclusions. The assumption of independence of observations was confirmed. Although the clinical psychologist had administered multiple assessments to some clients, no client had more than one of their assessments in the database. Multicollinearity, the unacceptably high correlation between two or more predictors, was assessed using the VIF (variance inflation factor) and tolerance statistics; a VIF > 10 or a tolerance less than .2 or greater than .8 may indicate collinear variables (Field, 2009). However, no predictor exceeded a VIF of 10 or fell outside the allowable tolerance range, indicating the absence of multicollinearity. The Durbin-Watson statistics for each regression model fell between 1.5 and 2.5, indicating the independence of errors (Tabachnick & Fidell, 2007). The assumption of homoscedasticity was also met. This was measured by checking the plots of standardized residuals, which were approximately equal in all values of the dependent variables.

Normality was assessed using Skewness and Kurtosis statistics as well as the Shapiro-Wilks statistic. All variables showed univariate normality according to Garson's (2009) cutoffs of -2 to +2 for Skewness and Kurtosis, with the exception of Auditory Attention, which was leptokurtic (10.83). The Shapiro-Wilk test compares the scores in the sample to a normally distributed sample with the same mean and standard deviation; significance ($p < .05$) indicates the sample is statistically significantly different from the normal distribution. Thus, nonsignificance ($p > .05$) indicates the sample close to a normal distribution (Field, 2009). Using Shapiro-Wilk's statistic Visual Auditory Learning, Concept Formation, Visual Matching, General Information, Auditory Attention, Reading Fluency, Passage Comprehension, Oral Comprehension, Pair Cancellation, and the two response and attention quotients significantly differed ($p < .05$) from the normal distribution. Numbers Reversed, Decision Speed, Memory for Words, and Math Fluency did not significantly differ ($p >.05$) from the normal distribution.

Although many of the variables were not normally distributed according to the Shapiro-Wilk statistic, this can be expected in clinical populations. Many of the individuals in the sample are on the outskirts of the normal distribution, likely relating to why they are a clinical population. A meta-analysis by Frazier et al. (2004) found up to 20-point differences in cognitive performance between ADHD and non-ADHD populations. However, the assumption of normality is robust with large sample sizes (i.e., more than 15 cases per predictor; Garson, 2009). Since few predictors were used in the regression analysis, this condition was satisfied.

Z-scores were used to test for univariate outliers, whereas the Mahalanobis statistic, the Leverage statistic, and Cook's distance were used to test for multivariate outliers and influential observations. Using Z-scores, a cut-off of 3.29 was established to find potential outliers (Field, 2009). Eight potential univariate outliers emerged throughout the data. Two of these cases accounted for six of the eight univariate outliers. Further, these two cases were the only two that exceeded the Mahalanobis critical value of $X^2(13) = 27.688$, $p < .01$, with values of 40.98 and 28.16. Cook's distance greater than 1 are cause for concern in regression (Cook and Weisberg, 1982). One case, with a Mahalanobis statistic of 40.98, had a Cook's distance above one (1.366), indicating it influenced the model as a whole. Leverage values measure undue influence on the regression coefficients; Stevens (2002) sets a cut-off point of $3\ (k + 1)/n$, where k is the number of predictors in the model, and *n* is the sample size. The cut-off for this analysis was .265 (3(13+1)/158); only one case, the same problematic case discussed above, was near the cut-off (.261). For the above reasons, the two cases were removed from the analyses, bringing the sample to $n = 156$.

The results of the originally planned analyses are presented first, followed by the results of the re-specified analyses. The original analyses failed in the initial stage, in which the WJ III data and the IVA data were individually tested for model fit. The software application could not converge on a proper solution. This was evidenced by paths (standardized regression weights) exceeding one, by only three significant paths in the model out of the possible ten that were estimate, by negative error variances, and by several near zero correlations. The path from the Auditory Response Quotient to Auditory Consistency, as reported by a standardized regression weight, was $r^2 = 1.176$, the path from the Visual Response Quotient to Visual Consistency was $r^2 = 2.625$, and the path from the Auditory Attention Quotient to Auditory Vigilance was $r^2 = 3.455$. Since, standardized regression weights can only range from -1 to +1, the paths were inadmissible.

The paths from the Full-Scale Attention Quotient to the Visual Attention Quotient, from the Visual Attention Quotient to Visual Focus, and from the Visual Attention Quotient to Visual Speed were significant ($p < .01$) while the seven other estimated paths were non-significant. A high proportion of non-significant paths indicate that the model does not explain the data well. The error variances of Auditory Vigilance ($R^2 = -12720.95$), Auditory Consistence ($R^2 = -166.322$), and Visual Consistency ($R^2 = -2261.65$) were found to be negative. Variances range from 0 to 1 and represent squared correlations; since a squared value cannot have a negative solution these variances are mathematically impossible and inadmissible. Finally, there are eight near zero and three greater than one squared multiple correlations in the model, which are the explained variances (R^2) of the variables. Table 1 contains the squared multiple correlations of the IVA's variables. For these reasons a solution could not be converged upon, the data could not be fit to the model, and the analyses were halted.

For the first re-specified hypothesis, at the Stratum I, the test level, bivariate correlations were run between the thirteen WJ III subtests (independent variables) and the two IVA quotients (dependent variables). Variables with significant correlations ($p < .05$) or that were near significance ($p < .10$) were considered in the regression equations. Table 2 contains bivariate correlations of the variables kept for the regression analysis. Two significant correlations and three that was near significance were found. Visual Matching ($r = .137$, $p < .10$), Math Fluency ($r = .137$, $p < .10$), and Decision Speed ($r = .166$, $p < .05$) correlated with the Full-Scale Response Quotient and were kept for the regression analysis. Pair Cancellation ($r = .135$, $p < .10$) and Math Fluency ($r = .257$, $p < .001$) correlated with the Full-Scale Attention Quotient and were kept for the regression analysis.

Using a stepwise regression, only Decision Speed ($B = .298$, SE = .128, $p < .05$) was found to predict the Full-Scale Response Quotient ($R^2 = .034$, F (1, 156) = 5.444, $p < .05$). This means that Decision Speed accounted for 3.4% of the variance in the Full-Scale Response Quotient, leaving 96.6% of the variance unaccounted for. Visual Matching and Math Fluency were dropped from the model because they did not account for any significant proportion of the variance over and above Decision Speed.

Using a second stepwise regression, only Math Fluency ($B = .556$, SE = .162, $p < .001$) was found to predicted the Full-Scale Attention Quotient ($R^2 = .073$, F (1, 156) = 12.262, p

< .001). This means that Math Fluency accounted for 7.3% of the variance in the Full-Scale Attention Quotient, while leaving 92.7% of the variance unaccounted for. Pair Cancellation was dropped from the model because it did not account for any significant proportion of the variance over and above Math Fluency.

For hypothesis 2, at Stratum II, the broad cognitive ability level, bivariate correlations were run between Short-Term Memory, Processing Speed, and Fluid Intelligence and the IVA's two Full-Scale Quotients. One significant correlation emerged between Processing Speed and the Full-Scale Response Quotient ($r = .169$, $p < .05$) was kept for the regression analysis. Using a stepwise regression, Processing Speed ($B = .283$, SE = .133, $p < .05$) was found to predict the Full-Scale Response Quotient ($R^2 = .028$, F (1, 154) = 4.510, $p < .05$). This means that Processing Speed accounted for 2.8% of the variance in the Full-Scale Response Quotient, while leaving 97.2% of the variance unaccounted for.

Hypothesis 3, at Stratum III, bivariate correlations were run between general intellectual ability and the two Full-Scale Quotients. It was found that general intellectual ability did not significantly correlated with the Full-Scale Response Quotient ($r = .014$, $p > .05$) or with the Full-Scale Attention Quotient ($r = .032$, $p > .05$). Thus, regressions were not run for this hypothesis.

Since few significant correlations or correlations that were near significance were found in the *a priori* hypotheses, it was decided to run *post hoc* correlations between the subtest-level hypothesized predictors (H_1) and the two Auditory and Visual scale scores that comprise each of the two Full-Scale Quotients. It could be that a predictor correlates to either the Auditory scale or the Visual score, but not both. If this was the case, a correlation on one of the scale scores could be masked in the Full-Scale Quotient by a highly non-significant relationship to the other scale score. The results of the bivariate correlations are summarized in Table 3. There was one significant correlation (Math Fluency; $p < .001$) and six correlations that were near significance (Visual Matching, Decision Speed, Math Fluency [twice], Spatial Relations, and Auditory Attention; $p < .05$). As per general guidelines in *post hoc* analyses, p-values were constrained to help control for Type I error.

Discussion

Attention Deficit Hyperactivity Disorder is assessed using various modalities such as behavioural observations, respondent checklists, cognitive ability scores, and computerized performance tests. The use of cognitive ability tests to help diagnose ADHD has recently gained popularity (Frazier, et al., 2004). However, diagnoses of ADHD using tests of cognitive ability, like the Woodcock-Johnson III, have primarily relied on mean differences between individuals with and without ADHD. Until the current investigation, only three discriminant validity studies had explored the relationship between the WJ III and ADHD. All three studies were not peer-reviewed and appeared in one of two WJ III technical or training manuals. The present investigation sought to determine whether the WJ III has an attention factor by factor analyzing WJ III and IVA+Plus®'s scores. However, the software application could not converge on a proper solution and analyses were modified.

Yed's Law states that when a problem or scenario has two equally expected, but antithetic solutions, an unpredicted third, intermediate option will emerge as the appropriate solution (Yednoroz, 2010). Such is the case with the statistical analyses of the current investigation. Finding a convergent solution, with good model fit would have been optimal.

Conversely, not finding model fit combined with the inability to conduct another type of data analysis would have been a waste of time. Although either result seemed likely to occur, a third, unlikely solution surfaced: the model did not converge and the analyses were modified to correlation and regression.

Separate analyses were run between the IVA's attention quotient and response quotient, and thirteen of the WJ III's subtests (H_1), three of the broad cognitive abilities (H_2), and general intellectual ability (H_3). Visual Matching, Math Fluency, Decision Speed, and Pair Cancellation may be related to the IVA's quotient scores, which are the IVA's primary diagnostic categories (Sandford & Turner, 1995). Although these six subtests related to measures of ADHD, Math Fluency and Decision Speed were found to predict a small amount of the variation in the IVA's two primary measures of ADHD. Additionally, Visual Matching, Writing Fluency, Spatial Relations, Retrieval Fluency, and Auditory Attention were found to be correlated to the IVA's variables in the *post hoc* analyses. It could be that these subtests are related to ADHD.

Both the *a priori* and *post hoc* findings should be interpreted with caution, especially the *post hoc* findings. With the exception of Math Fluency, which had one significant correlation ($p < .001$), all other correlations were near significance ($p < .05$) in the *post hoc* analyses. Given the large number of hypothesized relationships and the relatively few significant relationships found, the question of chance enters into the discussion; were these findings spurious?

The *a priori* hypotheses were based on a wide array of empirical research findings (i.e. Wechsler study comparisons, executive functioning literature, and three WJ III and ADHD studies). Each of the variables in hypothesis 1 and 2 were backed by evidence from at least two previous studies (Assesmany et al., 2001; Barkley, Grodzinsky, & DuPaul, 1992; Barkley, 1997;

Burns, et al.,, 2009; Floyd et al., 2003; Floyd et al., 2006; Ford et al., 2003; Frazier et al., 2004; Lahey, et al., 1988; Lerner & Yasutake, 2001; Phelps et al., 2005; Schwean & Saklofske, 1998; Vesley, 2001). Given the empirical backing, more significant correlations were expected to have been found. Hypotheses 1 specified 26 potential correlations, while only two were significant; hypothesis 2 specified six potential correlations, while only one was significant; hypothesis 3 specified two potential correlations and no significant correlations were found.

Two possible explanations for this arise, one from the measures and one from the sample. Both the WJ III subtests and the IVA require multiple cognitive abilities to perform any given subtest and to complete the IVA. For instance, Pair Cancellation requires visual processing, fine-motor regulation, and processing speed to complete the task. To an extent, the IVA requires these same abilities, along with sustained attention and divided attention. It could be that those tests that were found to significantly predict the IVA's scores most closely relate to the same set of abilities needed to perform the IVA.

Alternatively, the sample's composition could have caused the lack of significance and the inability to converge on a solution for the IVA's factor structure. Although both adults and children may display similar strengths and weaknesses as a result of their ADHD, they are differentiated by their level coping abilities and learning strategies. By adulthood, an individual with ADHD has presumably learned effective coping and learning strategies to help in their daily functioning. These strategies and mechanisms are likely to affect their testing scores, thus, it is suggested that the analyses be run on samples containing only children or only adults.

Math Fluency significantly predicted the Full-Scale Attention Quotient. Math Fluency is a timed test measuring the rapid calculation of basic math problems. Its relationship to the attention quotient could stem from the testing method, which requires sustained attention for a fixed period of time, something the attention quotient also purports to measure. However, if this were the case, we would expect Writing Fluency, Retrieval Fluency, and Pair Cancellation, to name a few, to also correlate with the attention quotient. Math Fluency was the only math-based subtest in the analyses. One explanation could be that completing a time-based, numerical calculation task draws on particular executive or working memory processes that are weak in those with ADHD, whereas time-based language tasks may rely on separate processes that are not as strongly affected by ADHD. However, the ADHD literature consistently shows that math deficits in individuals with ADHD are uncommon and that only having a deficit in math is even more uncommon.

These findings could be a function of the sample not being representative of the population. The Math Fluency's mean was 83.57 and Calculation's mean was 88.37 ($p < .001$), however, the expected mean in a normal population is 100. This indicates that Math Fluency is about 1.5 standard deviations what is expected and Calculation is about 1 standard deviation below what is expected. Although math deficits can occur in ADHD populations, its presence without the presence of other deficit areas (e.g. in Writing Fluency or Reading Fluency) is highly unlikely. It could be that the anomalous occurrence of a math deficit, without any other deficit areas in this sample caused Math Fluency to be able to predict the presence of ADHD via the IVA quotients. Thus, this finding requires further testing before it can be generalized.

Decision Speed significantly predicted the Full-Scale Response Quotient and then correlated with the Auditory Response score in the *post hoc* analyses. Decision speed has been cited as being deficit in individuals with ADHD (e.g. Barkley et al. 1990), which could result from attentional, focus, or executive functioning issues. Like Math Fluency, Decision Speed requires the timed, mental manipulation of numbers. This lends further support to the working hypothesis as a viable explanation of the findings as opposed to being explained by chance. Use a 2x2x2 ANOVA design, future research should test students with and without ADHD in math- and language-based areas on timed and untimed tasks to uncover whether timed, math-based tasks are better predictors of, or deficits in those with ADHD. Alternatively, using structural equation modeling techniques, a multi-trait multi-method design (MTMM) could be employed to determine whether the various WJ III testing methods (e.g. timed versus untimed tests) are predictive of ADHD.

The *post hoc* analyses provide guarded support for the hypotheses and the literature. The *p*-value was constrained to compensate for the chance of finding spurious results that are not replicable. Given the fact that only one significant relationship ($p < .001$) and five relationships that were near significance ($p < .05$) were found out of 52 possible relationships, caution should be yielded as the likelihood of finding at least 1 correlation out of 52 is high.

T-tests were used to assess differences between the removed cases and those that were kept. The proportion of males and females did not significantly differ between the two groups, nor did the average general intellectual ability. However, the average grade between the two groups did, such that those who were removed (grade 13) were an average 2.2 grades above those who were kept (grade 10.8). Although grades 0 through 11.9 are strong indicators of age (i.e. 0 years old to 17 years old), grades 12 and above are not. Throughout childhood and adolescence an individual's grade generally changes in a linear fashion. Once in adulthood, the grade-level ceases to increase once the individual has finished their schooling. Therefore, a grade of 12 or higher generally indicates the individual is over the age of 18, but not by how much. The grade difference between the two groups indicates that a higher proportion of the kept cases were still in elementary or high school, while the removed cases had a higher proportion of adults. This was not unexpected as the database was gathered from a school psychologist who also owns a private practice.

Five of the WJ III's subtests accounted for the majority of the missing values from the removed cases. All five subtests were from WJ III ACH. Since the WJ III's ACH subtests are most often used for educational programming, it is apparent why these five tests were primarily administered to those cases that were kept. The psychologist revealed that while all of the cognitive tests specified in the WJ III's model are generally administered to clients, fewer achievement tests are administered unless required.

Limitations and Future Opportunities

Evidence from the failed original analyses indicated that the data did not converge on the IVA's empirically specified model. Two competing explanations can be inferred, either the data is too dissimilar from the IVA's norming population to converge or fit the model or the model is poorly specified and does not exist as specified by its creators. The first explanation is difficult to conclude because the IVA's technical manual does not provide enough information to compare with. It only provides a general description of the norming group, without mention of any specific group characteristics. However, the authors do specify that the normative data were collected from normal individuals, without psychological disorder or attentional issues. The sample used in this investigation was clinical, clearly differing from the norming population−actually, the populations are polar opposites given the exclusion criteria for the norming population was one of the inclusion criteria for this clinical sample. Based on this, it can be concluded that one of the reasons for the data not converging on a solution and fitting the model is due to the difference between the sample used in this study and the norming population.

Consider what the test is used for and whom it is typically administered to. It is advertised to aid in diagnosing ADHD and is administered to those with attentional issues and whom are suspected of having ADHD−those in or soon to be in a clinical population. An analogy: adult clinical psychologists are not supposed to advertise to council children because it is outside of their competency, or outside of the "norming population" that they were trained on. I am left wondering why the IVA's norming population does not include those individuals whom the test is designed for−people with attentional problems. The norming sample is not representative of the population as a whole and is upwardly biased because the lower end of the attention spectrum was excluded. A further investigation of this is warranted.

Further complicating this issue is the IVA's great discriminant validity, correctly identifying 92% of the ADHD children and 90% of the non-ADHD children (Buros Institute, 2007). If the test is that accurate and is utilized only to supplement other diagnostic measures, does the poor factor structure and questionable norming population matter? That philosophical debate will be left for the reader to decide. If the test continues to show high discriminant validity and continues to be used on those with attentional issues, then a factor structure should have emerged in the present analyses. The IVA is 92% effective in identifying ADHD, so the data should fit the IVA's factor structure and the full scale quotients should have had higher correlations to the WJ III variables, as backed by the literature. This elicits more questions about the IVA, than answers the research questions.

The results of both the planned analyses and modified analyses mandate a concurrent validation of the IVA to uncover its validity in detecting various facets of attention and ADHD. The reason that the IVA has not been broadly accepted in practice could be due to its validity or lack thereof. The Buros Institute' review of the IVA condemns its validity, citing research on it is "limited" and that its predictive validity in differentiating between the four ADHD subtypes is "unsubstantiated" (Buros Institute, 2007). Is measuring a person's ability to click or to refrain from clicking a mouse button when presented with visual or auditory stimuli an appropriate method of diagnosing ADHD?

A closer look at the IVA's stamina variable may help explain some of the criticism surrounding the IVA's validity. The explained variance for both visual stamina and auditory stamina were near zero. Also, the Visual Stamina (r = -.057) and Auditory Stamina (r = .010) scores had near zero correlations with the latent variables. Stamina compares the mean reaction/decision times of correct responses during the first 200 trials and the last 200 trials. It purports to measure the ability to sustain attention and effort over time. However, the stamina score measures the slope, not the intercept. It does not control for the variation in the initial reaction time; it only measures the raw difference between the start and finish reaction times. It is likely that those with ADHD have slower initial reaction times than those without because of their deficit processing speed. It could be the initial reaction times are better at predicting ADHD than the change in it. Latent growth modeling techniques could be used to examine whether the intercept and/or the slope of the stamina scores better predictor ADHD. Although this measurement issue does not condemn the stamina scores, it does cast doubt on its validity.

I would also like to illuminate the incongruence between stamina's definition and its operationalization. It proposes to help "identify problems related to sustaining attention and effort over time" (Sandford & Turner, 1995). The problem lies in purporting to measure sustained effort over time. Since effort is a subjective feeling, it is highly unlikely that a measure of change in reaction time captures the construct of effort. Implying that effort is being measured is incorrect and could be insulting to an individual with ADHD. Stamina score are likely to decrease over time for all individuals, indicating a drop in "sustained attention and effort." However, individuals with ADHD will need to increase their effort to sustain attention, even if their reaction time decreases. A poor stamina score could be psychologically damaging to those who put forth an abundance of effort but are unable to maintain their reaction time because of their ADHD. It is suggested that either the stamina score be operationalized to include a subjective measure of effort or effort be taken out of its definition.

Yed's Law also prevailed in the choice of outcome measure. When compiling the database, behavioural measures of attention and of ADHD would have been preferred, considering the practice standards of diagnosing ADHD (Pliszka & AACAP Work Group on Quality Issues, 2007); however, they were not available. CPTs are intended to help diagnose ADHD, generally supplementing ADHD assessments to help rule out alternative disorders. Use of the IVA in the analyses was an unexpected middle-ground between having no measure of ADHD available and using behavioural checklists.

The IVA's questionable diagnostic utility, the inability to fit the data to its factor structure, the poor test-retest reliability, and the lack of available literature on the IVA's factor structure or norming population made it difficult to conclusively support or fail to support the hypotheses. It is suggested that similar regression analyses be conducted between the WJ III's variables and behavioural checklists. Given this, the true nature of the relationship between the WJ III and attention/ADHD might surface.

This original purpose of this paper was to identify whether the Woodcock-Johnson III had an attention factor, by conducting a cross-validation with the Integrated Visual and Auditory Continuous Performance Test. Since a convergent solution could not be found, the analyses shifted to correlation and regression. Although results from the re-specified analyses lacked a degree of novelty and only provided mild support for previous research, several areas of research in need of further study were illuminated–the most important of which detected a major possible flaw in a popular commercially-sold measure.

References

Abikoff, H., Courtney, M. E., Szaibel, P. J., & Koplewicz, H. S. (1996). The effects of auditory stimulation of the arithmetic performance of children with ADHD and nondisables children. *Journal of Learning Disabilities*, *29*, 238-246.

American Psychiatric Association. (2000). *Diagnostic and Statistical Manual of Mental Disorders (4th ed.).* Washington, DC: American Psychiatric Association.

Arbuckle, J.L. (2006). *AMOS users guide version 7.0.* United States of America: SPSS Inc.

Assesmany, A., McIntosh, D., Phelps, L., & Rizza, M. G. (2001). Discriminant validity of the WISC-III with children classified with ADHD. *Journal of Psychoeducational Assessment*, *19*, 137-147.

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

Barkley, R. A. (1998). *Attention-deficit hyperactivity disorder: A handbook for diagnosis and treatment (2nd ed.).* New York: Guilford.

Barkley, R., DuPaul, G., & McMurray, M. (1990). Comprehensive evaluation of attention deficit disorder with and without hyperactivity as defined by research criteria. *Journal of Consulting and Clinical Psychology*, *58*, 775-798.

Barkley, R., Grodzinsky, G., & DuPaul, G. (1992). Frontal lobe functions in attention deficit disorder with and without hyperactivity: A review and research report. *Journal of Abnormal Child Psychology*, *20*, 163-188.

Braden, J., & Alfonzo, V. (2003). The *WJ III Tests of Cognitive Abilities* in cognitive assessment courses. In F. Schrank, & D. Flanagan, *WJ III Clinical Use and Interpretations: Scientist-Practitioner Perspectives.* San DIego: Academic Press.

BrainTrain. (2010). *IVA+Plus Standard scale scores*. Retrieved 10 10, 2010, from BrainTrain: http://www.braintrain.com/professionals/adhdtesting/ivaplus_scoring.htm

Bridgett, D. J., & Walker, M. E. (2006). Intellectual functioning in adults with ADHD: A meta-analytic examination of full scale IQ differences between adults with and without ADHD. *Psychological Assessment*, *18*(1), 1-14.

Browne, M. W., & Cudeck, R. (1989). Single sample cross-validation indices for covariance structures. *Multivariate Behavioral Research, 24*, 445-455.

Burns, N., Nettelbeck, T., & McPherson, J. (2009). Attention and intelligence: A factor analytic study. *Journal of Individual Differences*, *30*, 44-57.

Buros Institute. (2007). *Test reviews online*. Retrieved 12 21, 2010, from Buros institute of mental measurements: http://buros.unl.edu/cgi-bin/BUROS/buros_display.cgi

Byrne, B. M. (2001). *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. Mahway, NJ: Lawrence Erlbaum Associates.

, C. L., & Tamm, L. (2000). Responsiveness of children with attention deficit-hyperactivity/disorder to reward and response cost: Differential impact on performance and motivation. *Journal of Consulting and Clinical Psychology*, *68*, 73-83.

Carleson, C. L., Mann, M., & Alexander, D. K. (2000). Effects of reward and response cost on the performance of motivation of children with ADHD. *Cognitive Therapy and Research*, *24*, 87-98.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor analytic studies.* New York:

Cambridge University Press.

Carroll, J. B. (1997). The three-stratum theory of cognitive abilities. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 122-130). New York: Guilford Press.

Carroll, J. B. (2003). The higher-stratum structure of cognitive abilities: Current evidence supports G and about ten broad factors. In H. Nyborg, *The scientific study of general intelligence: Tribute to Arthru R. Jensen* (pp. 5-22). Amsterdam: Pergamon Press.Cizek, G. J. (2003). Review of the *Woodcock-Johnson III*. In B. S. Plake, & J. C. Impara, *The fifteenth mental measurements yearbook* (pp. 1020-1024). Lincoln, NE: Buros Institute of Mental Measurement.

Cook, R. D., Weisberg, S. (1982). *Residuals and influence in regression*. New York: Chapman & Hall.

Cummings, J. (1995). Review of the *Woodcock-Johnson Psycho-Educational Battery- Revised*. In J. Conoley, & J. Impara, *The Twelfth mental measurements yearbook* (pp. 1113-1116). Lincoln, NE: Buros Institute of Mental Measurement.

Field, A. (2009). *Discovering Statistics Using SPSS* (Third ed.). London, Sage.

Flanagan, D. P., McGrew, K. S., & Ortiz, S. (2000). *The Wechsler Intelligence Scales and Gf-Gc Theory: A contemporary approach to interpretation.* Needham Heights, MA: Allyn & Bacon.

Floyd, R. G., Bergeron, R., & Alfonso, V. C. (2006). Cattell–Horn–Carroll cognitive ability profiles of poor comprehenders. Reading and Writing: An Interdisciplinary Journal, *19,* 427–456.

Floyd, R. G., Shaver, R. B., & McGrew, K. S. (2003). Interpretation of the Woodcock-Johnson III tests of cognitive abilities: Acting on evidence. In F. A. Schrank, & D. Flanagan, *WJ III clinical use and interpretation* (pp. 1-46, 403-408). New York: Academic Press.

Ford, L., Keith, T. Z., Floyd, R. G., Fields, C., & Schrank, F. A. (2003). Using the Woodcock-Johnson III tests of cognitive abilities with students with attention deficit/hyperactivity disorder. In F. A. Schrank, & D. P. Flanagan, *WJ III clinical use and interpretation: Scientist-practitioner perspectives* (pp. 319-344). San Diego, CA: Academic Press.

Ford, L., Percy, A., & Negreiros, J. (June, 2010). Canadian cognitive assessment training. *Paper presentated at the annual meeting of the Canadian Psychological Association, Winnipeg, Manitoba*.

Ford, L., Swart, S., Negreiros, J., Lacroix, S., & McGrew, K. (2010). *Woodcock-Johnson III NU* tests of cognitive abilities and tests of achievement with Canadian populations. *Woodcock-Johnson III Assessment Service Bulletin Number 12*, Rolling Meadows, IL: Riverside Publishing.

Frazier, T. W., Demaree, H. A., & Youngstrom, E. A. (2004). Meta-analysis of intellectual and neuropsychological test performance in attention-deficit/hyperactivity disorder. *Neuropsychology*, *18*(3), 543-555.

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

Garcia-Sanchez, C., Estevez Gonzalez, A., Suarez Romero, E., & Junque, C. (1997). Right hemisphere dysfunction in subjects with attention-deficit disorder with and without hyperactivity. *Journal of Child Neurology*, *12*, 107-115.

Garson, D. G. (2009, 7 13). *Statnotes: Topics in Multivariate Analysis*. Retrieved 10 28, 2009, from http://faculty.chass.ncsu.edu/garson/pa765/statnote.htm

Hibberland, D., & Saklofske, D. H. (1996). The *Wechsler Adult Intelligence Scale-Third Edition*. *Canadian Journal of School Psychology*, 74-76.

Hicks, P., & Bolan, L. (1996). Review of the *Woodcock-Johnson Psycho-Educational Battery-Revised*. *Journal of School Psychology*, *6,* 93-102.

Hoff, K. E., Doepke, K., & Landau, S. (2002). Best practice in the assessment of children with attention deficit/hyperactivity disorder: Linking assessment to intervention. In J. Grimes, & A. (. Thomas, *Best practices in school psychology* (pp. 1224-1251). Bethesda, MD: National Association of School Psychologists.

Horn, J. L., McArdle, J. J., & Mason, R. (1983). When is invariance not invariant: A practical scientist's look at the ethereal concept of factor invariance. *The Southern Psychologist, 1*, 179-188.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1-55.

Kamphaus, R. W., & Campbell, J. M. (Eds.). (2006). *Psychodiagnostic assessment of children: Dimensional and categorical approaches*. Hoboken: Wiley. Retrieved from PsycINFO.

Kamphaus, R. W., & Frick, P. J. (2005). *Clinical assessment of child and adolescent personality and behaviour.* New York, New York: Springer.

Kamphaus, R. W., Windsor, A. P., Rowe, E. W., & Kim, S. (2005). A history of intelligence test interpretation. In D. P. Flanagan, & P. L. Harrison, *Contemporary Intellectual Assessment* (pp. 23-38). New York: The Guilford Press.

Kessler, R. C., Adler, L. A., Barkley, R., & et al. (2005). Patterns and predictors of attention defi cit/hyperactivity disorder persistence into adulthood: Results from the National Comorbidity Survey Replication. *Biological Psychiatry*, *57*, 1442-1451.

Kessler, R. C., Adler, L., Barkley, R., & et al. (2006). The prevalence and correlates of Adult ADHD in the United States: results from the National Comorbidity Survey Replication. *American Journal of Psychiatry*, *163*, 716-723.

Krane, E., & Tannock, R. (2001). WISC-III third factor indexes learning problems but not Attention Deficit/Hyperactivity Disorder. *Journal of Attention Disorders*, *5*(2), 69-78.

Kronenberger, W., & Meyer, R. (1996). *The child clinician's handbook.* Boston, MA: Allyn and Bacon.

Lahey, B., Pelham, W., Schaughency, E. A., Atkins, M., Murphy , H., Hynd, G., et al. (1988). Dimensions and types of attention deficit disorder. *Journal of American Child and Adolescent Psychiatry*, *27*, 330-335.

Lee, S., & Stefany, E. (1995). Review of the *Woodcock-Johnson Psycho-Educational Battery-Revised*. In J. Conoly, & J. Impara, *The twelfth mental measurement yearbook* (pp. 1116-1117). Lincoln, NE: The University of Nebraska Press.

Lerner, J., & Yasutake, D. (2001). School-age AD/HD sample. In K. McGrew, & R. W. Woodcock, *Technical Manual. Woodcock-Johnson III.* Itasca, IL: Riverside Publishing.

Locke, S. R., McGrew, K. S., & Ford, L. (2011). A Multiple Group Confirmatory Factor Analysis of the Structural Invariance of the Cattell—Horn-Carroll Model of Cognitive Abilities Across Matched Canadian and US samples. The Woodcock-Muñoz Foundation Press Bulletins, pp. 1-34.

Mandell, D.S., Thompson, W. W., Weintraub, E. S., DeStefano, F., & Blank, M. B. (2005). Trends in diagnosing rates for Autism and ADHD at hospital discharge in the context of other psychiatric diagnoses. *Psychiatric Services*, *56*, 56-62.

Mather, N., & Woodcock, R.W. (2001). Woodcock-Johnson® III tests of cognitive abilities: Examiner's manual standard and eMaxtended batteries. Itasca, IL: Riverside.

McGrew, K. S. (1997). Analysis of the major intelligence batteries according to a proposed comnprehensive Gf-Gc framework. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds), *Contemporary intellectual assessment: Theories, tests, and issues (2nd Ed)* (pp. 151-179). New York: Guilford.

McGrew, K. S. (2005). The Cattell-Horn-Carroll theory of cognitive abilities: Past, present, and future. In D. P. Flanagan, & P. L. Harrison (Eds), *Contemporary intellectual assessment: Theories, tests, and issues, (2nd Ed)* (pp. 136-182). New York: Guilford.

McGrew, K. S. (2009). Editorial: CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*, *37*, 1-10.

McGrew, K. S., & Woodcock, R. W. (2001). *Technical Manual. Woodcock-Johnson III*. Itasca, IL: Riverside Publishing.

Phelps, L., McGrew, K., Knopik, S., & Ford, L. (2005). The general (g), broad, and narrow CHC stratum characteristics of the WJ III and WISC-III tests: A confirmatory cross-battery investigation. *School Psychology Quarterly*, *1*, 66-88.

Pliszka, S., & AACAPWork Group on Quality Issues. (2007). Practice parameter for the assessment and treatment of children and adolescents with attention-deficit/hyperactivity disorder. *Journal of the American Academy of Child and Adolescent Psychiatry, 46(7)*, 894–921. doi:10.1097/chi.0b013e318054e724.

Prifitera, A., & Dersh, J. (1993). Base rates of WISC-III diagnostic subtest patterns among normal, learning-disabled, and ADHD samples. *Journal of Psychoeducational Assessment, WISC-III Monograph*, 43-55.

Rapport, M. D., Chung, K., Shore, G., Denney, C. B., & Isaacs, P. (2000). Upgrading the science and technology of assessment and diagnosis: Laboratory and clinic-based assessment of children with ADHD. *Journal of Child Clinical Psychology*, *29*, 555-568.

Sanford, J., & Turner, A. (1994). *Manual for the Intermediate visual and auditory continuous performance test (version 2.6).* Richmond, VA: BrainTrain 727 Twin Ridge Lane.

Sandford, J., & Turner, A. (1995). *Intermediate Visual and Auditory Continuous Performance Test: Interpretation Manual.* Richmond, VA: BrainTrain.

Sandoval, J. (2003). Review of the *Woodcock-Johnson III.* In B. Plake, & J. Impara, *The fifteenth mental measurements yearbook* (pp. 1024-1027). Lincoln, NE: Buros Institute of Mental Measurements.

Schrank, F. A., & Flanagan, D. P. (2003). *WJ III clinical use and interpretation: Scientist-practitioner perspectives.* San Diego, CA: Academic Press.

Schrank, F. A., Flanagan, D. P., Woodcock, R. W., & Mascola, J. (2002). *Essentials of the WJ III tests of cognitive abilities assessment.* New York: John Wiley.

Schwean, V., & Saklofske, D. (1998). WICS-III assessment of children with attention deficit/hyperactivity disorder. In A. Prifitera, & D. Saklofske, *WISC-III clinical use and interpretations: Scientist-practitioners perscriptives* (pp. 91-118). San Diego, CA: Academic Press.

Tabachnick, B., & Fidell, L. (2007). *Using Multivariate Statistics* (5th ed.). New York: Pearson.

Taub, G. E., & McGrew, K. (2004). A confirmatory factor analysis of Cattell-Horn-Carroll theory and cross-age invariance of the *Woodcock-Johnson Tests of Cognitive Abilities* III. *School Psychology Quarterly*, *19*, 72-87.

Teeter, P., & Smith, P. L. (1993). WISC-III and WJ-R: prediction and discriminant validity for students with severe emotional disturbance. In B. Bracken, & S. McCallum, *Journal of Psychoeducational Assessment: Advances in psychoeducational assessment* (pp. 114-124). Brandon, VT: Clinical Psychology Publishing Co.

Vesley, B. (2001). Grades 1 through 6 normal, LD, AD/HD sample. In K. McGrew, & R. W. Woodcock, *Technical Manual. Woodcock-Johnson III.* Itasca, IL: Riverside Publishing.

Wielkiewicz, R. M. (1990). Interpreting low scores on the WISC-R third factor: It's more than disctactibility. *Psychological Assessment*, *2*(1), 91-97.

Wechsler, D. (1996). *Wechsler Intelligence Scale for Children-Third Edition, Canadian.* San Antonio, TX: The Psychological Corporation.

Wechsler, D. (2001). *Wechsler Intelligence Scale for Children-Third Edition, Canadian.* San Antonio, TX: The Psychological Corporation.

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children-Fourth Edition.* San

Antonio, TX: The Psychological Corporation.

Wechsler, D. (2004). *Wechsler Intelligence Scale for Children-Fourth Edition, Canadian.* San Antonio, TX: The Psychological Corporation.

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale-Fourth Edition, Canadian.* San Antonio, TX: The Psychological Corporation.

Woodcock, R. W. (1997). The *Woodcock-Johnson Tests of Cognitive Ability—Revised*. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison, *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 230-246). New York: Guilford.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III.* Itasca, IL: Riverside Publishing.

Woodcock, R. W., McGrew, K. S., Schrank, F. A., & Mather, N. (2001, 2007). *Woodcock-Johnson III Normative Update.* Rolling Meadows, IL: Riverside Publishing.

Yednoroz, C. (2010). On Yed's Law. *Personal Communication*. Retrieved, December 28, 2010, Windsor, Ontario.

Table 1

Square Multiple Correlations for the IVA's variables

Variable	R^2
Auditory Attention Quotient	.000
Visual Attention Quotient	.998
Visual Response Quotient	.741
Auditor Response Quotient	.024
Auditory Prudence	.075
Auditory Consistency	1.384***
Auditor Stamina	.000
Visual Prudence	.006
Visual Consistency	6.890***
Visual Stamina	.003
Visual Speed	.155
Visual Focus	.507
Visual Vigilance	.548
Auditory Speed	.014
Auditory Focus	.017
Auditory Vigilance	11.940***

Note: ***inadmissible solution.

Table 2

Preliminary Bivariate Correlations for Hypothesis 1.

Variables	*r*	*p*
Full-Scale Response Quotient by Visual Matching	.137	.087
Full-Scale Response Quotient by Decision Speed	.166*	.038
Full-Scale Response Quotient by Math Fluency	.137	.089
Full-Scale Attention Quotient by Math Fluency	.257**	.001
Full-Scale Attention Quotient by Pair Cancellation	.135	.092

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

Table 3

Post-hoc Bivariate Correlations

Variables	*r*	*p*
Auditory Response Quotient by Visual Matching	.189*	.018
Auditory Response Quotient by Decision Speed	.195*	.015
Auditory Response Quotient by Math Fluency	.197*	.014
Auditory Response Quotient by Spatial Relations	.171*	.033
Auditory Attention Quotient by Math Fluency	.323**	.001
Visual Attention Quotient by Auditory Attention	.162*	.043

Note: $*p < .05$, $**p < .001$.

Appendix A

Descriptions of WJ III NU COG Clusters and Tests Used in the Analyses

CHC Factor and Description	Test and Description
Comprehension-Knowledge (*Gc*): The depth and breadth of a person's acquired knowledge. This factor is analogous to the traditional notion of crystallized intelligence.	**Verbal Comprehension** (Test 1): Comprised of four subtests which together provide a measure of general language development, lexical knowledge and ability to apply this knowledge on verbal reasoning tasks (VBCMPAS).
	General Information (Test 11): A measure of general acquired (verbal) knowledge (GENINFAS).
Long-Term Retrieval (*Glr*): The ability to store and retrieve, often through association, information, concepts or facts fluently from memory.	**Visual-Auditory Learning** (Test 2): A paired-associative memory task that measures the ability to encode and retrieve visual-auditory symbolic information. A controlled learning task with corrective feedback (VALAS).
	Retrieval Fluency (Test 12): A set of three open-ended probes that measure the ability to fluently retrieve words within a specified limited period of time (RETFLUAS).
Visual-Spatial Thinking (*Gv*): The ability to store and recall visual stimuli and to synthesize, analyze, manipulate, and perceive visual patterns.	**Spatial Relations** (Test 3): A task requiring the ability to identify which two or three parts that, when combined, form a target visual figure (SPARELAS).
	Picture Recognition (Test 13) A measure of visual recognition memory of figures of common objects (PICRECAS).
Auditory Processing (*Ga*): The ability to discriminate, analyze, and synthesize auditory stimuli.	**Sound Blending** (Test 4): A measure of the ability to synthesize auditory stimuli (phonemes) (SNDBLNAS).
	Auditory Attention (Test 14): A measure of the ability to discriminate sounds in the presence of increasingly distracting auditory stimuli (AUDATNAS).

Fluid Reasoning (*Gf*): Problem-solving in relatively novel situations, particularly those requiring deductive and inductive thinking.	**Concept Formation** (Test 5): An inductive concept rule formation task that also requires mental flexibility. A controlled learning tasks with corrective feedback and reinforcement (CONFRMAS).
	Analysis Synthesis (Test 15): A mathematically-based deductive reasoning task that requires the application of rules from a key to the solving of logic problems. A controlled learning tasks with corrective feedback and reinforcement (ANLSYNAS).
Processing Speed (*Gs*): Speed of mental processing when performing relatively simple Cognitive tasks under conditions requiring sustained attention and concentration.	**Visual Matching** (Test 6): A task measuring the ability to rapidly discriminate and identify two identical numbers within a line of numbers (VISMATAS).
	Decision Speed (Test 16): A measure of the ability to rapidly identify pictures of two objects, from within a line of object pictures, that are conceptually related (DECSPDAS).
	Pair Cancellation (Test 20): A short, timed task designed to measure speed at which matching pairs of numbers can be found.
Short-Term Memory (*Gsm)*: The ability to consciously store, maintain, and use information presented within a few seconds.	**Numbers Reversed** (Test 7): A working memory task requiring the retention and mental manipulation of a sequence of numbers (NUMREVAS).
	Memory for Words (Test 17): A memory span test requiring the ability to retain and repeat a sequence of unrelated words (MEMWRDAS).

Appendix B

Descriptions of WJ III NU ACH Clusters and Tests Used in the Analyses

CHC Factor and Description	Test and Description
Reading (*Grw*): The ability to identify and comprehend written language.	**Letter-Word Identification** (Test 1): A test requiring the identification and pronunciation of printed letters and words; sight word recognition is used (LWDINTAS).
	Reading Fluency (Test 2): A measure of the ability to read printed statements rapidly and respond with either true of false (RDGFLAS).
	Passage Comprehension (Test 9): A task requiring the identification of a missing key word that makes sense in the context of a written passage (PSGCMPAS).
Math (*Gq*): The ability to manipulate symbols and to reason procedurally with quantitative information and relations.	**Calculation** (Test 5): A test measuring mathematical ability by performing various mathematical calculations (CALCAS).
	Math Fluency (Test 6): A measure calculation speed, by adding, subtracting, and multiplying rapidly (MTHFLUAS).
	Applied Problems (Test 10): Analyzing and solving orally presented, practical mathematical problems (APPROBAS).
Written Language (*Grw):* The ability to write meaningfully while applying lexical, grammatical, and syntactical rules.	**Spelling** (Test 7): Spelling letter combinations that are regular patterns in written English (SPELLAS).
	Writing Fluency (Test 8): Formulating and writing simple sentences rapidly (WRTFLUAS).
	Writing Samples (Test 11): Writing meaningful sentences for a given purpose (WRTSMPAS).
Oral Language: The ability to understand and comprehend the English Language orally.	**Oral Comprehension** (Test 15): Listening to a short passage and providing the missing final word (ORLCMPAS).

Appendix C

DSM-IV Diagnostic Criteria for diagnosing Attention Deficit (Hyperactivity) Disorder:

A. Either (1) or (2)

1) Six or more of the following symptoms of **inattention** have persisted for at least six months to a degree that is maladaptive and inconsistent with the developmental level:

Inattention

- Often fails to give close attention to details or makes careless mistakes in schoolwork, work, or other activities
- Often has difficulty sustaining attention in tasks or play activities
- Often does not seem to listen when spoken to directly
- Often does not follow through on instructions and fails to finish schoolwork, chores, or duties in the workplace (not due to oppositional behaviour or failure of comprehension)
- Often has difficulty organizing tasks and activities
- Often avoids, dislikes, or is reluctant to engage in tasks that require sustained mental effort (such as schoolwork or homework)
- Often loses things necessary for tasks or activities at school or at home (e.g. toys, pencils, books, assignments)
- Is often easily distracted by extraneous stimuli
- Is often forgetful in daily activities

2) Six or more of the following symptoms of **hyperactivity-impulsivity** have persisted for at least 6 months to a degree that is maladaptive and inconsistent with the developmental level:

Hyperactivity

- Often fidgets with hands or feet or squirms in seat
- Often leaves seat in classroom or in other situations in which remaining seated is expected
- Often runs about or climbs excessively in situations in which it is inappropriate (in adolescents or adults, may be limited to subjective feelings of restlessness)
- Often has difficulty playing or engaging in leisure activities quietly
- Often talks excessively
- Is often “on the go” or often acts as if “driven by a motor”

Impulsivity

- Often has difficulty awaiting turn in games or group situations
- Often blurts out answers to questions before they have been completed
- Often interrupts or intrudes on others, e.g. butts into other children's games

B. Some hyperactivity - impulsive or inattentive symptoms that cause impairment were present before the age of 7 years.

C. Some impairment from the symptoms is present in more than two or more settings (e.g. at school or work or at home).

D. There must be clear evidence of clinically significant impairment in social, academic, or occupational functioning.

E. The symptoms do not occur exclusively during the course of a Pervasive Developmental Disorder, Schizophrenia, or other Psychotic Disorder, and are not better accounted for by another mental disorder (e.g. Mood Disorder, Anxiety Disorder, Dissociative Disorder, or a Personality Disorder).

Based on these criteria, three types of ADHD are identified:

- ADHD, Combined Type: if both criteria 1A and 1B are met for the past 6 months
- ADHD, Predominantly Inattentive Type: if criterion 1A is met but criterion 1B is not met for the past six months
- ADHD, Predominantly Hyperactive-Impulsive Type: if Criterion 1B is met but Criterion 1A is not met for the past six months.

Vita Auctoris

Mr. Sean Robert Locke was born in 1987 in Windsor, Ontario, Canada. After completing grades one through three at King Edward and grades four through eight at John Campbell, he graduated high school from the Hon. Vincent Massey in Windsor. He completed his undergraduate was the University of Windsor in Honours Psychology and Criminology in 2009. Throughout his undergraduate and Master's tenure he was a volleyball coach (for six years), a youth leader (for four years), a mentor to some, and a SSHRC recipient. In 2011, he will be going to the University of Saskatchewan for his Ph.D. in Kinesiology and to be supervised to become an educational psychologist. He likes long walks on the beach, and has an affection for Junior Mints® from time to time.