Cognitive Variability in High-Functioning Individuals & its Implications for the Practice of Clinical Neuropsychology

by

Eliyas Jeffay

A thesis submitted in conformity with the requirements for the degree of Master of Arts

Department of Psychology
University of Toronto

© Copyright by Eliyas Jeffay 2010
Cognitive Variability in High-Functioning Individuals & its Implications for the Practice of Clinical Neuropsychology

Eliyas Jeffay
Master of Arts
Department of Psychology
University of Toronto
2010

Abstract
Knowledge of the literature pertaining to patterns of performance in normal individuals is essential if we are to understand intraindividual variability in neurocognitive test performance in neuropsychiatric disorders. Twenty-five healthy individuals with a high-level of education were evaluated on a short neuropsychological battery which spanned several cognitive domains. Results indicated that cognitive abilities are not equally distributed within a sample of healthy, high-level functioning individuals. This may be of interest to neuropsychologists who might base clinical inference about the presence of cerebral dysfunction, at least in part, on marked variation in a patient’s level of cognitive test performance. The practice of deductive reasoning in clinical neuropsychology may be prone to false-positive conclusions about cognitive functioning in neuropsychiatric disorders where base-rates of cognitive impairments are low and pre-existing educational achievements are high.
Acknowledgments

I was fortunate to have the support of some very special people who helped nurture and better my work.

I would like to thank Dr. Konstantine Zakzanis for his constant insight, interest, and most importantly, availability throughout the entire MA project. Truly, you are an inspiration to me. From the bottom of my heart, thank you for everything. I also wish to express my sincere appreciation to my subsidiary advisor, Dr. David Nussbaum, whose experience and expertise of the clinical field aided this project tremendously. Lastly, I would like to acknowledge the aid and assistance of Dr. Zakzanis’ lab; not only do I call this place my new ‘home’ but my colleagues as my friends.

I dedicate this thesis to my parents, Abbas Ali and Leila Jeffay, who instilled upon me the importance of education and the value of determination and hard-work. I am forever indebted to their constant guidance, encouragement and love.
# Table of Contents

Acknowledgments .......................................................................................................................... iii  
Table of Contents ........................................................................................................................... iv  
List of Tables .................................................................................................................................. vi  
List of Figures ................................................................................................................................. vii  

Chapter 1 Literature Review ........................................................................................................... 1  
1 The Best Performance Method ................................................................................................... 2  
2 The Limited Sensitivity of Neuropsychological Tests ............................................................... 4  
3 Base Rates of Impairment in Clinical Neuropsychology .......................................................... 6  
4 Cognitive Variability .................................................................................................................. 8  
  4.1 Importance of Intraindividual Variability ......................................................................... 10  
  4.2 Intelligence as a Source of Intraindividual Variability ..................................................... 11  

Chapter 2 Methods ........................................................................................................................ 24  
1 Participants ............................................................................................................................... 14  
2 Materials ................................................................................................................................... 15  

3 Description of Measures ........................................................................................................... 16  
  3.1 Trail Making Test ............................................................................................................. 16  
  3.2 Wide Range Achievement Test - Fourth Edition - Reading ............................................. 17  
  3.3 Ruff 2&7 Selective Attention Test ................................................................................... 17  
  3.4 Rey-Osterrieth Complex Figure Test .............................................................................. 17  
  3.5 Digit Span ....................................................................................................................... 17  
  3.6 Controlled Oral Word Association Test .......................................................................... 18  
  3.7 Wisconsin Card Sorting Task ......................................................................................... 18  
  3.8 Judgment of Line Orientation – Form V ......................................................................... 18
List of Tables

Table 1. Demographic characteristics of the high functioning participants .................. 15

Table 2. Neuropsychological measures used and the cognitive domains they represent ........ 16

Table 3. Means and standard deviations of raw scores for each neuropsychological test measure .......... 24
List of Figures

Figure 1. A comparison of the means of each neuropsychological test measure across all participants converted into z-score ........................................................................................................... 27

Figure 2. Mean variances of the neuropsychological test measures across all participants converted to z-scores .................................................................................................................. 29

Figure 3. The mean variance of each cognitive domain across all participant.................. 30

Figure 4. Maximum discrepancy and adjusted maximum discrepancy scores by participant for all neuropsychological measures........................................................................................................ 31

Figure 5. Frequency distribution of maximum discrepancy and adjusted maximum discrepancy scores.................................................................................................................. 32

Figure 6. Scatterplot depicting each participant’s FSIQ (in SD units) as a function of the mean of their mean difference from FSIQ on all test measures. The difference scores increased negatively with better FSIQ scores (Pearson’s r = -0.75).................. 33

Figure 7. Scatterplot depicting each participant’s maximum discrepancy (in SD units) as a function of the mean of their z-scores on all test measures. Intraindividual variability decreased with better overall cognitive test performance (Pearson’s r = -0.35).................................................................................................................. 34
Chapter 1  
Literature Review

The basic element of test score analysis within the context of a neuropsychological examination is a significant discrepancy between any two or more scores. Marked quantitative discrepancies in a person’s performance suggest that some abnormal condition is interfering with that person’s overall ability to perform at their characteristic level of cognitive functioning (Lezak, Howieson & Loring, 2004; Silverstein, 1987). Statistically, a significant discrepancy between any two or more scores inherently implies an unlikely difference between a pair of variables (such as neuropsychological test scores) compared to typical individuals. The probability of obtaining such a discrepancy by chance or measurement error is low (e.g., p < .05) if the “true” difference between the scores is zero (Matarazzo & Herman, 1984). As Schretlen, Munro, Anthony and Pearlson (2003) so elegantly note:

The underlying assumption of this approach to the assessment of intraindividual variability is that the person’s “true” abilities ($a_1 = a_2$), as measured by the test score pair, are identical. The null hypothesis in this case could be expressed as $H_0: a_1 = a_2$. By extension, when a battery of neuropsychological tests measuring i abilities is administered, and all possible pairs of scores are compared, the null hypothesis is that all of the “true” scores are identical (i.e., $H_0: a_1 = a_2 = ...a_i$). (p. 864)

As these authors demonstrate in their study on the range of intraindividual variability in neuropsychological test performance, even among normal, healthy persons, there is no reason to accept this null hypothesis as correct. That is, both the complexity of the human central nervous system and individual differences in the organization of neural circuitry on which various mental abilities depend argue against the likelihood that any individual will be endowed with identical levels of ability across all domains of cognitive functioning (Hawkins & Tulsky, 2003; Schretlen et al., 2003; Schretlen, Testa, Winicki, Pearlson & Gordon, 2008).

Knowledge of the literature pertaining to patterns of performance in normal individuals is essential as due consideration is not always given to normal intraindividual variability in the context of interpreting neuropsychological test performance. To do so is important for several
reasons. First, a clinician will often employ the “Best Performance Method” (see Lezak et al., 2004) when trying to interpret neuropsychological test performance. This approach may be an inappropriate heuristic of pre-injury cognitive functioning where intraindividual variability exists. Second, it has been demonstrated that a large number of tests with little to no sensitivity to, for example, mild Traumatic Brain Injury (mTBI) and various other neuropsychiatric disorders, are often employed for clinical and research purposes (see Zakzanis, Leach, & Kaplan, 1999). Lastly, in the instance of low base rates of cognitive impairment, it may be that the neuropsychological examination reveals little beyond that of normal intraindividual variability rather than evidence of impairment secondary to brain injury or psychiatric disorder. What follows is a description of each of these tenets as they pertain to the practice of clinical neuropsychology along with a description of cognitive variability. Furthermore, illustrations of how clinicians may not always give due consideration to normal intraindividual variability in neuropsychological test performance are cited to clarify why this may be important if we are to place diagnostic practice in clinical neuropsychology on firmer scientific grounds.

1  The Best Performance Method

In the absence of premorbid neuropsychological data, Lezak et al. (2004) proposed the Best Performance Method as a means to estimate premorbid levels of cognitive function in an individual within the context of a neuropsychological examination. The individual’s “best performance – whether it be the highest score or set of scores, non-scorable behaviour not necessarily observed in a formal testing situation, or evidence of premorbid achievement – serves as the best estimate of premorbid ability” (Lezak et al., 2004, p. 97). The method is based on the assumption that “...given reasonably normal conditions of physical and mental development, there is one performance level that best represents each person’s cognitive abilities and skills generally” (Lezak et al., 2004, p. 97). Within the context of a neuropsychological examination, when utilizing this approach, the clinician will deduce that impaired cognition exists when scores stand 1.5 standard deviations (SD) from the highest obtained score (Lezak et al., 2004).

Clinicians and researchers have been critical regarding the cut-off criterion of 1.5 SD when utilizing the Best Performance Method. For example, it has been argued that a high degree of scatter is to be expected in normal healthy individuals and, therefore, a criterion of 1.5 SD could be potentially misleading (Franzen, Burgess & Smith-Seemiller, 1997). As well, it has been
shown that some cognitive abilities naturally decline throughout the life-span. Thus, an aging population may achieve scores using the Best Performance Method that is suggestive of a mild to moderate decline even in the absence of disease or cognitive decline (Mortensen, Gade, & Reinisch, 1991). A further criticism as it relates to the utility of the Best Performance Method is that it ignores the law of regression to the mean and maximizes the premorbid IQ estimate based on an extreme chance fluctuation or intraindividual variability. For example, this method of estimating premorbid intelligence has been demonstrated to almost always overestimate premorbid IQ by about 12 IQ points when extrapolating from Wechsler Adult Intelligence Scale – Third Edition (WAIS-III; The Psychological Corporation, 1997; Wechsler, 1997) samples (Reynolds, 1997). Indeed, evidence demonstrating variable performance patterns on the WAIS-III scales indicates that there is not a one-to-one correspondence between all types of cognitive abilities examined by this specific measure (Wechsler, 1999).

It may also be that the Best Performance Method is limited with respect to its ability to predict premorbid function based on cognitive variability research in normal, healthy individuals. For example, Heaton, Grant and Matthews (1991), Heaton, Miller, Taylor and Grant (2004) and Iverson, Brooks and Holdnack (2008) computed the likelihood of abnormality in normal, healthy adults based on various definitions of abnormality. When defining abnormality as a score that is greater than one standard deviation from the mean, they determined that 10%-15% of the total number of test scores in a neuropsychological test battery with more than 20 measures would be deemed abnormal by chance alone. They found that these abnormalities were not entirely corrected for when increasing the definition of abnormality to scores greater than 2 standard deviations from the mean. Additionally, these researchers suggested a direct proportional relationship with the number of tests within a battery and the number of abnormal test scores to be expected from a population of healthy, normal adults. Namely, as the number of tests increased, the chances of finding an abnormal score also increased.

Using the patient’s highest performance level on a single scale or subtest as representing overall cognitive ability may also be an oversimplification of more varied interrelationships among intellectual abilities (Lynch & McCaffrey, 1997). For example, several examinations (Christensen et al., 1999; Gyrke, Prifitera, & Sharp, 1991; Hultsch, MacDonald, Hunter, Levy-Bencheton & Strauss, 2000; Kaufman, 1976; Matarazzo, Daniel, Prifitera & Herman, 1988; Matarazzo & Prifitera, 1989; McLean, Kaufman & Reynolds, 1989) of the performance patterns
of healthy, neurologically intact individuals comprising the standardization samples of the various Wechsler Intelligence Scales revealed that significant score variation across the subtests of these measures is the norm, not the exception. For reasons such as these, the Best Performance Method is recommended only for cases of moderate to severe brain injury, although still yet, practitioners have often used this approach in the interpretation of neuropsychological test scores following mTBI (Mortensen et al., 1991; Schoenberg, Duff, Scott, & Adams, 2003) and other conditions (see Zakzanis et al, 1999).

2 The Limited Sensitivity of Neuropsychological Tests

The term sensitivity within the context of a neuropsychological examination refers to a neuropsychological test measures’ ability to accurately detect a given abnormality. In the same context, specificity refers to the measures’ ability to accurately differentiate those with a certain abnormality from those with other or no abnormalities (Lezak et al., 2004).

The limited sensitivity and thus utility of many neuropsychological tests have been demonstrated (Binder, Iverson, & Brooks, 2009; Heaton et al., 1991, 2004). For example, in a comprehensive quantitative review of the research literature on mTBI, Binder, Rohling and Larrabee (1997) found small effect sizes (i.e., magnitude of differences between normal and impaired cases; see Zakzanis, 1998; 2001). These findings suggest that the maximum prevalence of persistent neuropsychological deficit after three months post-injury is likely to be small and as such, the neuropsychological assessment is likely to have positive predictive value of less than 50% (Binder et al., 1997). Consequently, these authors concluded that clinicians will more likely be correct when not diagnosing brain injury than when diagnosing a brain injury particularly in cases with chronic disability after mTBI.

Cicerone and Azulay (2002) found sensitivity levels of various neuropsychological test measures of attention to be low to non-existent in patients with post-concussion syndrome (PCS), a condition occurring in patients with persisting neurological complaints after the expected recovery following a traumatic brain injury. Thirty-two patients with documented PCS and their matched controls completed 6 neuropsychological measures of attention in order to assess their sensitivity, specificity, along with positive and negative predictive values in the detection of PCS. Based on their results, the specificities of all the tests were found to be adequate, however the sensitivities of the Stroop Test (ST; Golden, 1978), Trails Making Test Parts A & B (TMT;
Reitan & Wolfson, 1992), Ruff 2 & 7 Selective Attention Test (RSAT; Ruff & Allen, 1996) and Wechsler Adult Intelligence Scale – Fourth Edition (WAIS-IV; Wechsler, Coalson, & Raiford, 2008) Digit Span (DS; Wechsler et al., 2008) were all found to be very low. The authors concluded that inter-test variability may be due to test characteristics rather than individual characteristics.

Other examples illustrating the limited sensitivity of neuropsychological tests exist. For example, Zakzanis et al. (1999) found small effect sizes by way of a comprehensive quantitative review of the neuropsychological research across a broad range of neuropsychiatric disorders including Major Depressive Disorder (MDD), obsessive-compulsive disorder and schizophrenia, suggesting that base rates of cognitive impairments (or the sensitivity of neuropsychological tests) across neuropsychiatric disorders are also low.

Duff, Hobson, Beglinger and O’Bryant (2010) have also shown that the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS; Randolph, 1998) demonstrated excellent specificity but moderate to poor sensitivity to patients with Mild Cognitive Impairment (MCI). In their study, two groups, patients with MCI ($n = 72$, $M = 82.1$yrs, $SD = 6.9$yrs) and cognitively intact peers ($n = 71$, $M = 76.3$yrs, $SD = 6.9$yrs), completed the RBANS and objective cognitive deficits in memory, executive functioning, language and attention/processing speed were examined. Their results illustrated equivocal support for the use of the RBANS in mildly impaired individuals. Their data suggested that although older patients with MCI scored significantly below their cognitively intact peers, the sensitivity values for the indices on the RBANS were quite poor. Subsequently, the researchers concluded that based on the insensitivity to MCI, caution should be used when utilizing the RBANS in diagnosing a patient with MCI.

A further example of the limited sensitivity of neuropsychological tests can be found in the Attention Deficit-Hyperactivity Disorder (ADHD) literature. To this end, McGee, Clark and Symons (2000) examined the Conner’s Continuous Performance Task (Conner’s CPT), a very popular screening test for ADHD (Barkley, 1991; McGee et al., 2000). These authors wanted to determine if the test could distinguish between children with ADHD and other disorders relative to clinical controls. The authors found small effect sizes for the overall index of the Conner’s CPT suggesting that the test did not distinguish between children with ADHD and clinical controls. Additionally, the authors indicated that only 52% of the children with ADHD were
detected as having ADHD by the Conner’s CPT (limited sensitivity). This finding was consistent with the work of Halperin et al. (1990) who also found that the clinical utility of the Conner’s CPT as a diagnostic instrument in detecting ADHD was questionable.

Overall, an ideal neuropsychological test is one with an optimal balance of specificity and sensitivity (Lezak et al., 2004). Based on the research literature, the specificity of many neuropsychological test measures seem to be sufficient in their ability to differentiate patients with cognitive abnormality from those with other or no abnormalities (Amato & Denney, 2008; Duff et al., 2010; Henry, Merten, Wolf, & Harth, 2010; Passetti, Clark, Mental, Joyce, & King, 2008; Singhal, Green, Ashaye, Shankar, & Gill, 2009). In contrast, the ability to detect particular abnormalities (i.e., sensitivity) appears to be low to modest at best across many disorders. The low sensitivity of various neuropsychological test measures is explained by some researchers have attributed it to either the particular clinical condition studied (Duff et al., 2010), definition of abnormality (Duff et al., 2010; Etherton, Bianchini, Greve, & Heinly, 2005) or methodology used to study specificity and sensitivity (Dearth et al., 2005; Rogers, 1997). In sum, the clinical utility of many neuropsychological tests remains uncertain with regards to the accuracy of detecting a particular abnormality (i.e., its sensitivity).

3 Base Rates of Impairment in Clinical Neuropsychology

Base rates are defined as the probability of obtaining abnormal scores on a particular neuropsychological test measured as a percentage of the total population (Lezak et al., 2004). These probabilities are derived from a combination of specific experimental conditions, sample population(s) utilized and neuropsychological test measure(s) used (Ostrosky-Solis, Lopez-Arango, & Ardila, 2000). The ability of a neuropsychological test’s sensitivity and specificity to correctly classify, for example, a patient with mTBI versus those without mTBI depends on the base rate (Mitruschina, Boone, Razani, & D’Elia, 2005; Strauss, Sherman, & Spreen, 2006). Traditionally, the base rates of patterns of low scores have allowed clinicians to infer neuroanatomical abnormality (Lezak et al., 2004). Binder et al. (2009) demonstrated that when interpreting multiple test scores, obtaining low and high test score scatter may not be indicative of an abnormal pathology, but rather that of normal intraindividual variability. These authors suggested that proper understanding of abnormal function must be referenced against scatter of
test scores in normal, healthy samples. In this light, several studies have examined the score profiles of normal, healthy individuals in hopes of better understanding abnormality. The work of these authors is summarized below.

Economou (1999) studied the differences between immediate and delayed recall and working memory in 322 mid-age to old-age ($M = 64.59$, $SD = 9.27$) normal, healthy community dwellers. The participants were given the Logical Memory I (LM I), Logical Memory II (LM II) and the Letter-Number Sequencing (LNS) tasks from the Wechsler Memory Scale – Third Edition (WMS-III; Wechsler, 1997) in a single session. The results indicated that discrepancy between recall and working memory varied as a function of both age and education and that delayed recall-working memory was stable and resistant to normal age-related memory decline. Interestingly, the results also suggested that the greatest (however modest) discrepancy scores between recall and working memory was evident in the older and more educated group. In other words, the results of this study suggested that within-individual variability with regards to memory increased as education increased.

Additionally, Schretlen et al. (2003) recruited 197 normal, healthy adult participants ($M = 55.1$, $SD = 19.1$) and examined them across a wide range of cognitive domains using 15 different neuropsychological tests. The scores were standardized on a $z$-scale and the difference between each participant’s maximum and minimum scores were computed as a measure of their maximum discrepancy. Within their study, the maximum discrepancy between the scores ranged from 1.6 - 6.1 SDs which were suggestive of abnormal test scores. The authors noted that 2% of the sample produced maximum discrepancy scores of less than 2 SDs whereas the remaining 98% of the sample produced “marked quantitative discrepancies” which suggested that some sort of “abnormal condition is interfering with their ability to perform at their characteristic level of cognitive functioning” (Lezak et al., 2004, pg. 153). When the cut off was increased to 3 SDs, 66% of the sample population continued to exhibit test scores with marked quantitative discrepancies. Interestingly, these researchers also noticed that a modest inverse correlation between IQ and intraindividual variability was observed. In other words, as IQ scores increased, maximum discrepancy scores (intraindividual variability) decreased (Schretlen et al., 2003).

Ram, Rabbit, Stollery, and Nesselroade (2005) studied the weekly intraindividual variability (inconsistency) of older adults ($M = 65.9$ years, $SD = 7.1$) on a letter search task measuring
reaction time over a period of 36 weeks. Using multivariate data analysis, the authors found no evidence of regular state-like patterns and attributed the weekly variability to random noise. This was in contrast to what they originally hypothesized as they expected a significant amount of variability in the data collection which would remain relatively stable across weeks (Ram et al., 2005). Interestingly, upon further analysis, the authors discovered that the magnitude of random noise within each subject was negatively correlated with level of intelligence. In other words, as the level of intelligence increased, intraindividual variability decreased. These results were consistent with previous research conducted by Li and Lindenberger (1999). The findings of this study and that of Economou (1999) are however, in contrast to that found by Schretlen et al. (2003).

As a follow-up to their original study in 2003, Schretlen et al., (2008) studied the rates of abnormal test performance of different neuropsychological batteries of varied length, using three different cutoffs to define abnormality on 327 ($M = 54.8, SD = 18.8$) reasonably healthy adults. They found that the abnormal scores obtained by the neurologically healthy, normal adults were not entirely due to chance but rather a combination of age, sex, race, education and estimated premorbid IQ. The authors note that even after adjusting the scores to demographic norms, the number of abnormal scores obtained did not decrease significantly. Additionally, their findings suggested that as the number of tests within the battery increased, the chances of obtaining an abnormal score (based on any cut off level) also increased. This relationship has also been reported by Heaton et al. (1991, 2004) and Iverson et al. (2008).

4 Cognitive Variability

Normal cognitive variability might be an important factor to consider when determining if a particular test score or patterns of test scores are inferred as abnormal. Accordingly, what follows is a review of this literature.

The study of intraindividual variability is the study of the dynamic changes of how, when and why an individual, unconsciously, changes their behaviour or cognition over time to ultimately understand human behaviour (Baltes, Reese, & Nesselroade, 1977; Nesselroade & Ram, 2004). In this paradigm, individuals are not static beings with fixed cognitions and behaviours. Instead it assumes that any given individual has complex assortments of characteristics which vary at different rates (Nesselroade & Ram, 2004). Some characteristics may vary rapidly such as every
hour or day, whereas others may be more static and take as long as days, weeks or even years to vary and thus are relatively stable over long periods of time. The complex collection of varying and stable characteristics has been deemed as a “behavioural signature” which may uniquely identify any given individual – like a fingerprint (Shoda, Mischel, & Wright, 1994).

Specific subtypes of variability have also been proposed. Hultsch, MacDonald and Dixon (2002) classified variability into three subtypes; diversity, dispersion and inconsistency. Although others have also classified variability within a neuropsychological paradigm (Butler, Hokanson, & Flynn, 1994; Li & Lindenberger, 1999; Nesselroade & Ford, 1985; Nesselroade & Ram, 2004; Shammi, Bosman, & Stuss, 1998; Shoda et al., 1994; Slifkin & Newell, 1998), the definitions used by Hultsch et al. (2002) have been widely cited and generally accepted within the scientific community.

When evaluating a single task on a single occasion and comparing the results between persons, this variability is referred to as interindividual differences, or diversity. An example of diversity can be illustrated when evaluating and comparing scores on the Vocabulary subtest in the WAIS-III to measure differences between participants. The scores achieved by each individual are compared and contrasted across individuals. Evaluating within person variability along multiple measures on one single occasion is referred to intraindividual variability; or more specifically, dispersion. An example of this would be if an individual were tested on a single occasion on the Neuropsychological Assessment Battery (NAB; Stern & White, 2003), which contains multiple subtests to measure a wide array of cognitive skills and functions in adults. In this instance, the scores from different sub-tests within the battery would be analyzed for inconsistencies. Lastly, measuring within person variability along one single test on multiple occasions is also labeled as intraindividual variability but termed more specifically as, inconsistency (Stuss, Pogue, Buckle, & Bondar, 1994). For instance, this can be demonstrated when administering the Finger Tapping test (FTT; Reitan & Wolfson, 1993), which indexes motor speed and control. One could administer this measure to a patient with mTBI on multiple occasions throughout the recovery period in order to index recovery of motor speed and control over time.

Lastly, intraindividual change is defined as within person change that is enduring (Ram et al., 2005) and that can be caused by development, learning or aging (Nesselroade, 1991). This change can be a measure of cognitive or perceptual abilities or other nervous system activities.
For example, as individuals progress further into their development, they learn and experience new information, increasing their fluid intelligence ($G_f$) factor over the years. This change in $G_f$ from the early stages of human development to the early-to-mid stages of adulthood is considered intraindividual change. Similarly, the gradual decline in cognitive abilities from the late 20s until death also constitutes intraindividual change (Salthouse, 2009).

In sum, interindividual variability is defined as variation of specific abilities between persons on a single occasion of testing. Intraindividual variability is modeled as short term and temporary cognitive fluctuation which occurs within an individual. Intraindividual change however, is considered to be enduring and long-lasting change which occurs through learning, development and/or aging.

4.1 Importance of Intraindividual Variability

Intraindividual variability in cognitive performance was once considered meaningless, random errors (van Zomeren & Brouwer, 1987). Binder et al. (2009) stated that intuitively, since large test score scatter along with large discrepancy scores are common among healthy adults, a certain number of low test scores are to be expected. Additionally, measurement error, individual weaknesses in certain cognitive areas, motivation to complete, broadly defined psychological interference and fatigue increase likelihood of obtaining a low test score and potentially a misdiagnosis of abnormality within a clinical setting (Binder et al., 2009; Lezak et al., 2004; Mitrushina et al., 2006; Strauss et al., 2005).

It is argued that measures of central tendency provide statistical variance cushioning to add to the generalizability of the data. Although statistical measures of central tendency do offer invaluable information in an efficient manner, the uniqueness of the particular individuals’ performance scores is lost. Moreover, the totality of response patterns is overlooked when using methods of central tendency which, although seemingly irregular and random, may inherently have a logical and structural basis (van Zomeren & Brouwer, 1987). In this light, a growing body of literature has deemed intraindividual variability in cognitive performance as useful and as an informative method to distinguish individuals at risk for Alzheimer’s Disease (Caselli, Iaboli, & Nichelli, 2009), MCI, adults with arthritis (Hultsch et al., 2000), borderline personality disorder (BPD; Beblo et al., 2006), traumatic brain injuries (TBIs; Burton, Hultsch, Strauss, & Hunter, 2002), Parkinson’s disease (PD; Camicioli, Wieler, De Frias, & Martin, 2008; Frias, Dixon, Fisher,
Camicioli, 2007), attention hyper-activity disorder (ADHD; Klein, Wendling, Huettner, Ruder, & Peper, 2006; Zeeuw et al., 2008), schizophrenia (Reichenberg et al., 2006), malingering (Strauss, MacDonald, Hunter, Moll, & Hultsch, 2002) and even death\(^1\) (MacDonald, Hultsch, & Dixon, 2008).

### 4.2 Intelligence as a Source of Intraindividual Variability

Finally, intelligence as a factor of intraindividual variability has also been examined, albeit with equivocal results. Based on empirical studies, some researchers have concluded that participants with higher intelligence scores (broadly defined and measured) express more intraindividual variability from a neuropsychological standpoint as compared to those with lower intelligence scores (Binder et al., 2009; Hawkins & Tulsky, 2003; Iverson, Brooks, White, & Stern, 2008; Matarazzo & Herman, 1985; Matarazzo & Prifitera, 1989; Schretlen et al., 2008; Tulsky, Rolphus, & Zhu, 2000; Wechsler, 1997; Wechsler, 2008). Other researchers have however found exactly the opposite, stating that those with higher intelligence scores exhibit less intraindividual variability compared with those with lower intelligence scores (Hultsch et al., 2002; Li, Aggen, Nesselroade, & Baltes, 2001; Li & Lindenberger, 1999; Rabbitt, Osman, Moore, & Stollery, 2001; Ram et al., 2005). This discrepancy in findings may reflect the different test measures used in these empirical studies. Some researchers employed the WAIS-III Full Scale IQ (FSIQ) as their index of general intelligence whereas others have employed neuropsychological tests that were specifically intended to measure \(G_t\) or crystalized intelligence (\(G_c\)) or even both.

Many studies examining the relationship between intraindividual variability and intelligence have utilized sample populations which were of average or below average intelligence (as measured by VIQ, PIQ and FSIQ). As a result, some researchers have concluded that, theoretically, as intelligence increases, intraindividual variability increases whereas others have hypothesized the opposite relationship. To date however, researchers have only inferred what moderating role intelligence might play in terms of intraindividual variability. That is, no

\(^1\) This study examined the relationship between rate and inconsistency of speed performance on neuropsychological measures in combination of traditional accuracy-based markers of cognitive performance in an older sample. The authors concluded that inconsistency in speed performance was an early sign of neurological impairment and ultimately impending death.
experimental studies have been undertaken to systematically and objectively examine intraindividual cognitive variability in highly intelligence persons so as to confirm (or reject) the hypothesis that intelligence does play a moderating role in terms of neuropsychological test variability.

5 Purpose of the Present Study

The purpose of this study was to determine how much intraindividual variability would be evinced on a reasonably comprehensive battery of neuropsychological tests in a sample of highly intelligent healthy individuals. Since evidence of intraindividual neuropsychological test variability in healthy samples with “average” educational achievement (e.g., high school) is abundant (see Heaton, Ryan, Grant, & Matthews, 1996; Heaton et al., 2004; Palmer, Boone, Lesser & Wohl, 1998; Schretlen et al., 2003; 2008) and there is a discordance within the literature with regards to the relationship between intraindividual variability and intelligence, this study was particularly interested in determining how much intraindividual variability would be evinced in persons with a particularly high level of educational achievement. Accordingly, healthy individuals with a particularly high level of educational achievement were examined to determine if such neuropsychological test score variability was also characteristic in this unique sample of subjects. As previously noted, conventional clinical neuropsychology often bases clinical inferences about the presence of cerebral dysfunction, at least in part, on marked variation in a patient’s level of cognitive test performance (e.g., the Best Performance Method). This practice of deductive reasoning may be exceptionally prone to false positive conclusions about cognitive functioning in neurobehavioral and neuropsychiatric disorders where base rates of cognitive impairments are low and pre-existing educational achievements are high. Thus, we set out to address the following questions:

1. Do individuals with particularly high levels of educational achievement demonstrate variability in neuropsychological test performance?

2. Which (if any) cognitive domains are most variable in persons with a particularly high level of educational achievement?

3. Is the Best Performance Method an appropriate method to estimate pre-morbid intelligence in persons with a particularly high level of educational achievement?
4. Do normal, healthy, highly intelligent persons evince test scores in the abnormal range? In other words, is it possible to be highly intelligent yet neuropsychologically impaired?
Chapter 2
Methods

1 Participants

Our sample consisted of 25 graduate students in various departments of the University of Toronto (9 male, 16 female; ages 23 – 32 years, \( M = 25, SD = 2.27 \)). Participants were recruited via campus email. All participants were considered to be highly functioning individuals with a history of extensive education, defined as the attainment or near-attainment of the doctoral Ph.D. degree (i.e., no less than a Master’s Degree or a Doctorate degree in progress). Thus, the average education level varied from 16 to 26 years (\( M = 19.6, SD = 3.24 \)). English was the first language for the majority of the participants. Eight participants indicated that English was not their first language but were comfortable and fluent in English. To this end, all participants scored greater than an 11.9 grade reading level on the Wide Range Achievement Test 4, Reading subtest (WRAT4-Reading; Wilkinson & Robertson, 2006). Demographics for the sample are presented in Table 1 along with their means and SDs. Any participant with a known history of neurological disease, psychiatric illness, head injury, or stroke was excluded. Medical and psychiatric history for each participant was unremarkable in that there was no history of learning disorder, substance abuse or dependence, psychiatric disease or traumatic brain injury defined by way of having lost consciousness for any duration. None of the participants were currently using any psychotropic medications. Financial compensation was provided as an incentive to their participation in this study in the form of a random lottery with remuneration ranging from $5 to $100. All participants had equal chance-odds of winning any of the prizes.
Table 1.

Demographic characteristics of the high functioning participants.

<table>
<thead>
<tr>
<th></th>
<th>Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>25.4 (2.27)</td>
</tr>
<tr>
<td>Gender (% female)</td>
<td>64%</td>
</tr>
<tr>
<td>Education (years)</td>
<td>19.6 (3.24)</td>
</tr>
<tr>
<td>WRAT4-Reading (grade level)</td>
<td>12.78 (0.39)</td>
</tr>
<tr>
<td>Handedness (% right-handed)</td>
<td>88%</td>
</tr>
</tbody>
</table>

Note. WRAT4-Reading = Wide Range Achievement Test – Fourth Edition, Reading Trial.

Participants were treated in accordance with the Tri-Council Policy Statement in that informed consent was obtained and the participant was entirely aware that they could withdrawal their participation in this study at any given time while doing so without consequence (Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada, Social Sciences and Humanities Research Council of Canada, 1998). All participants were given a cover letter detailing the nature and purpose of the study along with how confidentiality was maintained. Participants were informed that their performance on cognitive tests was to be examined so as to further understand the range of possible test scores across a handful of popular clinical tests of memory and concentration as well as problem solving. To this end, they participated in pen-and-paper based neuropsychological tests which lasted approximately 90-minutes. All data were kept confidential with only the researchers associated with this study having access to the data files. The participants were directed to address all inquiries to the principle investigator.

2 Materials

A brief interview form was prepared for each participant to ascertain demographic variables (such as age, gender, years of education, first language and handedness) and to screen for any psychiatric, neurological or relevant medical history. A brief neuropsychological test battery consisting of the WRAT4-Reading subtest, TMT, RSAT, Rey Complex Figure Test (RCFT; Meyers & Meyers, 1995), DS, Controlled Oral Word Association Test (COWAT; Bechtoldt,
Benton, & Fogel, 1962), Wisconsin Card Sorting Test – 128 Card Version (WCST; Heaton, Chelune, Talley, Kay, & Curtiss, 1993), Judgment of Line Orientation Form-V (JLO; Benton, Hamsher, Varney, & Spreen, 1994), Grooved Pegboard Test (GPT; Lafayette Company, 2002), California Verbal Learning Test - II Standard Form (CVLT; Delis, Kramer, Kaplan, & Ober, 2000), and the Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 1997), including more specifically the, Vocabulary, Block Design, Similarities and Matrix Reasoning subtests, were used to index each participant’s level of cognitive functioning. These tests were chosen on the basis that they are some of the most commonly employed neuropsychological tests used by clinical neuropsychologists (Rabin, Borgos, & Saykin, 2008). Table 2 presents the measures employed in the study parceled by way of their respective cognitive construct domain(s).

Table 2.

Neuropsychological measures used and the cognitive domains they represent.

<table>
<thead>
<tr>
<th>Domain</th>
<th>Measure(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Orientation &amp; Attention</td>
<td>RSAT, TMT and DS</td>
</tr>
<tr>
<td>Visual Recognition</td>
<td>JLO</td>
</tr>
<tr>
<td>Verbal Memory</td>
<td>CVLT</td>
</tr>
<tr>
<td>Verbal Expression</td>
<td>COWAT and Vocabulary (from the WASI)</td>
</tr>
<tr>
<td>Construction</td>
<td>RCFT, Block Design (from the WASI)</td>
</tr>
<tr>
<td>Concept Formation</td>
<td>WCST, Matrix Reasoning and Similarities (from the WASI)</td>
</tr>
<tr>
<td>Motor Performance</td>
<td>GPT</td>
</tr>
</tbody>
</table>


3 Description of Measures

3.1 Trail Making Test

Participants were administered part A and part B of the TMT according to the guidelines presented by Strauss et al. (2006). When an error was made, the participant was notified and instructed to correct their mistake(s) and to continue until the test was completed. A sample trial
was presented prior to the test administration to insure adequacy of the instructions. Raw scores for the TMT were obtained by measuring the length of time required to complete the tasks (Reitan, 1979).

3.2 Wide Range Achievement Test - Fourth Edition - Reading

As per the WRAT4 manual (Wilkinson & Robertson, 2006), a stimulus sheet with 15 letters along with 55 words were presented to the participants individually which they were asked to read aloud. Raw scores for the WRAT4-Reading were obtained by summing the number of letters and words correctly pronounced.

3.3 Ruff 2&7 Selective Attention Test

As per the RSAT manual (Ruff & Allen, 1996), participants were asked to quickly and accurately scan and cross out 2’s and 7’s that were mixed in with other numbers and letters. The RSAT served as a measure of visual attention, both sustained and selective (Lezak et al., 2004). Raw scores for the RSAT were obtained by counting all the correct items in each trial and calculating the total speed and total accuracy of the scores as per the RSAT manual (Ruff & Allen, 1996).

3.4 Rey-Osterrieth Complex Figure Test

As per the RCFT manual (Meyers & Meyers, 1995), this test examines the ability to remember and recall a complex figure. Participants were asked to draw the complex figure on a white piece of paper for the Copy, Immediate Recall and Delayed Recall trials. Raw scores for the RCFT were obtained by comparing the subjects’ drawings to the stimulus figure and assessing which components had been included and how correctly they had been drawn (Meyers & Meyers, 1995).

3.5 Digit Span

Taken from the WAIS-IV, this subtest measures short term memory and attention by asking the participants to repeat an increasing length of digits read out loud to them (Lezak et al., 2004). The task consists of two trials; the forward trial, where the participant is asked to repeat in forwards directions and in the reverse trial, where they are to repeat in backwards order.
(Wechsler et al., 2008). The raw scores on the DS were calculated based on total correct responses as described in the WAIS-III Canadian technical manual (Wechsler, 2001).

3.6 Controlled Oral Word Association Test

This test was designed to measure verbal and semantic fluency. As per the original instructions (Bechtoldt et al., 1962), participants were asked to think of as many words as possible in 60 seconds following a phonemic (F, A, and S) or semantic (animals) cue. Exclusions for phonemic responses were proper nouns (such as Bob, Boston or Buick), numbers (such as four, three) and the repetition of a word with a different suffix (such as eat and eating). All responses were recorded by the psychometrist. The raw scores for the COWAT were obtained by summing the correct responses in the “F, “A” and “S” trials along with counting the total number of animals produced in the Animal trial (Bechtoldt et al., 1962).

3.7 Wisconsin Card Sorting Task

As per the instructions of the 128-card version (Heaton et al., 1993), this executive functioning test measures perseveration, set maintenance and learning. Participants were shown 4 stimulus cards and each stimulus card was distinct with respect to colour, form or number of items. They were instructed to match response cards from 2 decks of 64 cards and place it below any of the 4 stimulus cards. The sorting rules were not provided to the participants but individual feedback as to whether the card placement was correct or incorrect was given. After 10 consecutive correct card sorts, the sorting rule was changed to another principle. The first sorting principle was colour, then form followed by number of items. Testing was discontinued when the participants successfully sorted through the three sorting principles two times or when all the 128 cards had been used. The raw scores for the WCST were calculated based on the total number of errors, perseverative responses, perseverative errors, non-perseverative errors and number of categories completed (Heaton et al., 1993). Additionally, the percent scores for the previously stated categories in the WCST were calculated along with the number of sets that the participant failed to maintain (see Heaton et al., 1993).

3.8 Judgment of Line Orientation – Form V

As outlined by the manual (Benton et al., 1994), this test was designed to measure visual-spatial ability by assessing the ability to judge the orientation and angles of lines in space. Participants
were shown 30 pairs of angled lines and asked to match them along a response card. Subjects’ raw scores on the JLO were calculated based on total correct responses (Benton et al., 1994).

3.9 Grooved Pegboard Test

This task was designed to measure motor speed and accuracy (Lezak et al., 2004). Participants were presented with the groove pegboard instrument (Model 32025; Lafayette Instrument Company, PO Box 5729, Lafayette, IN, 47903, USA) which contained 25 rotated grooved-holes. Participants were instructed to fit the grooved pegs, one at a time, into the slotted holes angled in different directions within a 5x5 array. The participants performed the task twice, once with their dominant hand and once with their non-dominant hand (Lafayette Instrument Company, 2002). Participants’ raw scores on the GPT were calculated by measuring the length of time required to complete the task with the dominant and non-dominant hand respectively (Strauss et al., 2006).

3.10 California Verbal Learning Test – Second Edition

Following the instructions from the manual for the standard form (Delis et al., 2000), this task serves as a measure of learning, retrieval along with recognition. In the Learning trials of the task, the participants were asked to learn and recall words lists presented several times. Nine indices were used from this measure (List A Free Recall, List A Total Free Recall, List B Free Recall, Short-Delay Free Recall, Short-Delay Cued Recall, Long-Delay Free-Recall, Long-Delay Cued-Recall and Long-Delay Yes/No Recognition Hits). Raw scores for the CVLT were obtained by counting the number of correct items listed per trial, obtaining results for trial 1 and 5 of the List A Free Recall, total correct from trials 1-5 in the List A Free Recall, total List B Free Recall, total Short-Delay Free Recall, total Short-Delay Cued Recall, total Long-Delay Free Recall, total Long-Delay Cued Recall and total Long-Delay Yes/No Recognition (Delis et al., 2000).

3.11 Wechsler Abbreviated Scale of Intelligence

As per the manual (Wechsler, 1999), the complete WASI battery was used to measure each participant’s level of cognitive functioning with respect to vocabulary (Vocabulary subtest), verbal concept formation (Similarities subtest), visual-spatial construction (Block Design subtest) and logical reasoning (Matrix Reasoning subtest) (Lezak et al., 2004). Raw scores for
the four-subtest WASI were obtained by summing the total number of correct answers (Wechsler, 1999). A brief description of each subtest follows.

3.12 Vocabulary

In this subtest, the participants were shown a stimulus booklet with 42 words over 6 pages in list order increasing with difficulty. The participants were instructed to give brief definitions of each word after the psychometrist stated “what does this mean?” referring to the word in the trial. All responses were recorded in the WASI record booklet and each response was compared against the manual to determine a score of fully correct (2 points), partially correct (1 point) or incorrect (0 points). If the participants scored 4 consecutive scores of 0 or 4 scores of 0 out of 5 consecutive trials, the subtest was discontinued. Based on ages of the participants, the start point for all of the participants was set at word #9 (“bird”; Wechsler, 1997) and the end point was the very last word #42 (“panacea”). As per the manual, a practice trial was used prior to the administration of the test items to insure adequacy of the instructions.

3.13 Similarities

In this subtest, the participants were asked what each pair of words has in common. This subtest comprises of 26 pairs of words with increasing difficulty but based on the age of our sample, all participants started on pair #7 (“grapes and strawberries”) and ended with #26 (“law and freedom”; Wechsler, 1997). The psychometrist went through each pairs of words in the subtest and asked the participants “what does ___ and ___ mean?”. All responses were recorded in the WASI record booklet and each response was compared against the manual to determine if it was fully correct (2 points), partially correct (1 point) or incorrect (0 points). If the participants scored 4 consecutive scores of 0 or 4 scores of 0 out of 5 consecutive trials, the subtest was discontinued. As per the manual, a practice trial was used prior to the administration of the test items to insure adequacy of the instructions.

3.14 Block Design

In this subtest, the participants were asked to arrange a set of blocks (each block was red on 2 sides, white on 2 sides and half red and half white on 2 sides) according to a design pattern as a measure of their visual-spatial abilities. This subtest comprises of 13 increasingly difficult trials. Based on the age of our sample, all participants started on design #3 (Wechsler, 1997). The time
to complete each trial was recorded in the WASI record booklet and each response was compared against the manual to determine a score (up to 7 points). As per the manual, a practice trial was used prior to the administration of the test items to insure adequacy of the instructions.

3.15 Matrix Reasoning

In this subtest, the participants were shown a stimulus book with 35 increasingly difficult visual pattern completion patterns. The participants were instructed to choose from a multiple-choice array the item that best completed the pattern. Based on the age of our sample, the start question was #7 (Wechsler, 1997). No feedback was given and all responses were recorded in the WASI record booklet. A correct response was awarded 1 point and an incorrect response was awarded 0 points. If the participants scored 4 consecutive scores of 0 or 4 scores of 0 out of 5 consecutive trials, the subtest was discontinued. Two practice sample questions were used prior to the administration of the test items to insure adequacy of the instructions.

4 Neuropsychological Assessment (Procedure)

Each participant completed a clinical interview that was followed by the administration of the aforementioned test battery. Each test was administered in a standardized manner by a trained psychometrist under the supervision of a licensed clinical neuropsychologist. The order of the tests was as follows: WRAT4-Reading, COWAT, CVLT (Learning, List B, Short-Delay Free Recall and Short-Delay Cued Recall trials), RCFT-Copy trial, TMT, GPT, RCFT-Immediate Recall trial, JLO, DS, RSAT, WCST, CVLT (Long-Delay Free Recall, Long-Delay Cued Recall and Long-Delay Yes/No Recognition trials), RCFT (30-Minute Delayed trial) followed by the WASI (Vocabulary, Block Design, Similarities and Matrix Reasoning). The order was kept consistent for all participants. All tests were completed between 9:00 and 17:00 hrs at a time that was convenient for both the participant and the psychometrist.

For each test, raw scores were converted to standardized scores and compared to an appropriately selected demographic subset of normative data stratified by demographic variables made available by the source. This involved comparisons to normative data within the test manuals or, where appropriate, peer-reviewed, normative data studies. In terms of the latter, this was the case for the GPT, COWAT, JLO and TMT.
The normative data published by Bornstein (1985) were utilized (as described in Strauss et al., 2006, p. 960, Table A23.3) to calculate standardized scores for the GPT. The normative sample which Bornstein used was more applicable to this study’s current sample as Bornstein included a large, highly educated, right handed, Canadian sample who were recruited from college campuses and offices with the means stratified by age (Strauss et al., 2006).

The normative data published by Tombaugh, Kozak, and Reese (1999) were utilized (as described in Strauss et al., 2006, p. 751, Tables A11.36 & A11.37) to calculate standardized scores for the COWAT as they were more applicable to the current study’s sample population. Most other normative studies on verbal fluency and semantic fluency have looked at either a particular sample demographic (homosexual men, elderly, those with neurological disorders, ethnic groups…etc.) or they have used one of the many variants of the test (using CFL, PRW or BLT instead of FAS and using fruits, vegetables or something else instead of animals), all of which do not fit with the sample population in the current study. Tombaugh et al. (1999) used normal, healthy individuals with stringent recruitment criteria, large well described Canadian sample size and their data was separated by age and education. Also, the data from Tombaugh et al. (1999) are all taken from a Canadian sample (Carleton University – Ottawa, Ontario, Canada) whereas almost all the other normative studies use an American sample. Lastly, the authors use exactly the same procedures which this study used as original outlined by Bechtoldt et al. (1962).

With regards to the TMT, the normative data published by Bornstein (1985) were utilized (as described in Strauss et al., 2006, p. 633, Table A4.11) to calculate standardized scores as they were more applicable to the current study’s sample population based on nationality, educational level and lack of a history of neuropsychological disease.

The normative data published by Benton et al. (1994) were utilized (as described in Benton et al., 1994, p. 59, Table 5-3) to calculate standardized scores for the JLO as they were more applicable to the current study’s sample population based on educational level and age range. Additionally, as per the instructions of Benton et al., (1994) a total of 2 points were added to the raw scores of all female participants to correct for gender discrepancies (see Benton, Varney, & Hamsher, 1978).
The RCFT Copy, WCST Number of Categories Completed and WCST Failure to Maintain Set scores were excluded as little to no variability was offered by the conversion of their raw scores to $t$ or $z$-scores as indicated by their respective test manuals.

All statistical calculations were completed using the Statistical Package for the Social Sciences (SPSS) for Windows version 17.0 (SPSS Inc., 223 S. Wacker Drive, 11th Floor, Chicago, IL, 60606, USA).
Chapter 3
Results

The means and standard deviations of performance for each neuropsychological test across the sample are summarized in Table 3 for descriptive purposes. Our sample of 25 healthy high functioning participants demonstrated average to superior levels of cognitive functioning across the IQ measures employed in the study. The average estimate Full Scale IQ (FSIQ), by way of a WASI four-subtest prediction, was found to be 122.4 with a standard deviation of 10.4. This corresponds to 1.5 standard deviations above the defined average of 100. Additionally, it is important to note that the lowest FSIQ score was greater than this average at 102. Thus, it was seemingly reasonable to assert that the participants in this study were a rather uniform and unique sample of highly educated and high-functioning individuals.

Table 3.
Means and standard deviations of raw scores for each neuropsychological test measure.

<table>
<thead>
<tr>
<th>Test measure</th>
<th>n</th>
<th>M ± SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>COWAT (FAS Trial)</td>
<td>25</td>
<td>42.52 ± 10.04</td>
</tr>
<tr>
<td>COWAT (Animal Trial)</td>
<td>25</td>
<td>22.92 ± 5.16</td>
</tr>
<tr>
<td>TMT-A</td>
<td>25</td>
<td>30.49 ± 9.70</td>
</tr>
<tr>
<td>TMT-B</td>
<td>25</td>
<td>56.61 ± 22.67</td>
</tr>
<tr>
<td>DS</td>
<td>25</td>
<td>19.84 ± 3.21</td>
</tr>
<tr>
<td>JLO</td>
<td>25</td>
<td>26.24 ± 2.60</td>
</tr>
<tr>
<td>GPT Trial 1</td>
<td>25</td>
<td>57.94 ± 5.97</td>
</tr>
<tr>
<td>GPT Trial 2</td>
<td>25</td>
<td>64.94 ± 7.28</td>
</tr>
<tr>
<td>RSAT Speed</td>
<td>25</td>
<td>114.04 ± 21.87</td>
</tr>
<tr>
<td>RSAT Accuracy</td>
<td>25</td>
<td>118.16 ± 1.52</td>
</tr>
<tr>
<td>CVLT List A Trial 1</td>
<td>25</td>
<td>7.04 ± 1.67</td>
</tr>
<tr>
<td>CVLT List A Trial 5</td>
<td>25</td>
<td>13.52 ± 2.66</td>
</tr>
<tr>
<td>CVLT List A Total Trials</td>
<td>25</td>
<td>55.52 ± 8.68</td>
</tr>
<tr>
<td>CVLT List B</td>
<td>24</td>
<td>6.96 ± 2.05</td>
</tr>
<tr>
<td>CVLT Short Delay Free Recall</td>
<td>24</td>
<td>12.50 ± 3.02</td>
</tr>
<tr>
<td>CVLT Short Delay Cued Recall</td>
<td>24</td>
<td>12.67 ± 2.63</td>
</tr>
<tr>
<td>CVLT Long Delay Free Recall</td>
<td>24</td>
<td>12.71 ± 2.97</td>
</tr>
<tr>
<td>Test Description</td>
<td>N</td>
<td>Mean ± SD</td>
</tr>
<tr>
<td>------------------------------------------</td>
<td>---</td>
<td>-----------------</td>
</tr>
<tr>
<td>CVLT Long Delay Cued Recall</td>
<td>24</td>
<td>12.75 ± 2.80</td>
</tr>
<tr>
<td>CVLT Yes/No Recognition</td>
<td>24</td>
<td>15.12 ± 0.97</td>
</tr>
<tr>
<td>RCFT Immediate Recall</td>
<td>25</td>
<td>24.34 ± 5.17</td>
</tr>
<tr>
<td>RCFT Delayed Recall</td>
<td>25</td>
<td>24.28 ± 4.98</td>
</tr>
<tr>
<td>WCST Total Errors</td>
<td>21</td>
<td>16.90 ± 15.18</td>
</tr>
<tr>
<td>WCST Percent Errors</td>
<td>21</td>
<td>18.14 ± 10.63</td>
</tr>
<tr>
<td>WCST Perseverative Responses</td>
<td>21</td>
<td>13.19 ± 12.26</td>
</tr>
<tr>
<td>WCST Percent Perseverative Responses</td>
<td>21</td>
<td>14.24 ± 8.86</td>
</tr>
<tr>
<td>WSCT Perseverative Errors</td>
<td>21</td>
<td>11.24 ± 9.33</td>
</tr>
<tr>
<td>WCST Percent Perseverative Errors</td>
<td>21</td>
<td>12.26 ± 6.79</td>
</tr>
<tr>
<td>WSCT Non-Perseverative Errors</td>
<td>21</td>
<td>4.52 ± 3.12</td>
</tr>
<tr>
<td>WCST Percent Non-Perseverative Errors</td>
<td>21</td>
<td>5.14 ± 2.59</td>
</tr>
<tr>
<td>WASI Vocabulary</td>
<td>25</td>
<td>68.40 ± 6.02</td>
</tr>
<tr>
<td>WASI Matrix Reasoning</td>
<td>25</td>
<td>30.64 ± 2.34</td>
</tr>
<tr>
<td>WASI Similarities</td>
<td>25</td>
<td>42.28 ± 3.57</td>
</tr>
<tr>
<td>WASI Block Design</td>
<td>25</td>
<td>59.00 ± 11.95</td>
</tr>
<tr>
<td>WASI Verbal IQ</td>
<td>25</td>
<td>122.16 ± 10.50</td>
</tr>
<tr>
<td>WASI Performance IQ</td>
<td>25</td>
<td>117.76 ± 11.72</td>
</tr>
<tr>
<td>WASI Full-Scale IQ – 4 Subtest</td>
<td>25</td>
<td>122.36 ± 10.40</td>
</tr>
<tr>
<td>WASI Full-Scale IQ – 2 Subtest(^a)</td>
<td>25</td>
<td>121.56 ± 9.60</td>
</tr>
</tbody>
</table>

*Note.* RSAT = Ruff 2 & 7 Selective Attention Test, TMT = Trial Making Test (Trials A/B), DS = Digit Span (from Wechsler Adult Intelligence Scale – Fourth Edition), JLO = Judgment of Line Orientation, CVLT = California Verbal Learning Test - Second Edition, Standard Form, COWAT = Controlled Oral Word Association Test, WASI = Wechsler Abbreviated Scale of Intelligence, RCFT = Rey Complex Figure Task, WCST = Wisconsin Card Sorting Test, 128 Card Version, GPT = Grooved Pegboard Test, WRAT4-Reading = Wide Range Achievement Test – Fourth Edition, Reading Trial. Four subjects were not administered the WSCT as they were familiar with the sorting principles prior to the assessment. Additionally, one subject was not administered List B, Short Delay Free Recall, Short Delay Cued Recall, Long Delay Free Recall, Long Delay Cued Recall and the Yes/No Recognition Trial of the CVLT. \(^a\)Full-scale IQ – 2 Subtests is a standardized value calculated from the Vocabulary and Matrix Reasoning raw scores.

The raw means of the scores across all individuals were converted to a *z-score* and plotted in Figure 1. Since the Performance IQ (PIQ), Verbal IQ (VIQ) and FSIQ were calculated based on the various WASI subtests, they were excluded from Figure 1 to avoid the inclusion of
overlapping measures. In our sample, scores on all of the WASI subtests, RSAT and WCST Non-Perseverative Errors were found to be *High Average* to *Superior* across all participants (>0.6 z-score). Hereafter however, participants demonstrated marked variability in terms of their performances across the remaining test measures despite evidence of homogeneity in terms of FSIQ. It was found that 11 out of the 25 participants (44%) had at least one score that was classified as *Very Poor*, or “*Retarded*” (z-score < 2.0) as defined by Lezak et al. (2004). These findings are consistent in light of studies illustrating that IQ scores predict neurocognitive scores in persons of below average intelligence, but less so in individuals of above average intelligence (Diaz-Asper, Schretlen & Pearson, 2004).
Figure 1.

A comparison of the means of each neuropsychological test measure across all participants converted into z-score.
Note. RSAT = Ruff 2 & 7 Selective Attention Test, TMT = Trial Making Test (Trials A/B), DS = Digit Span (from Wechsler Adult Intelligence Scale – Fourth Edition), JLO = Judgment of Line Orientation, CVLT = California Verbal Learning Test - Second Edition, Standard Form, COWAT = Controlled Oral Word Association Test, WASI = Wechsler Abbreviated Scale of Intelligence, RCFT = Rey Complex Figure Task, WCST = Wisconsin Card Sorting Test, 128 Card Version, GPT = Grooved Pegboard Test.

The mean variances of the neuropsychological measures across all participants are illustrated in Figure 2. Marked variance (>1.0 SD) was found on the RCFT Delayed Trial, TMT-A, GPT Trial 1 and Trial 2, DS, WCST Percent Perseverative Responses, WCST Perseverative Errors, WCST Percent Perseverative Errors and all CVLT measures except for the Yes/No Recognition. Thus, marked variance was found on 16 out of the 33 (48.5%) measures employed.
Figure 2.

Mean variances of the neuropsychological test measures across all participants converted to z-scores.

Figure 3.

The mean variance of each cognitive domain across all participants.

**Note.** Verbal Expression = Controlled Oral Word Association Test and Vocabulary (from the Wechsler Abbreviated Scale of Intelligence), Construction = Rey Complex Figure Test and Block Design (from the Wechsler Abbreviated Scale of Intelligence), Orientation & Attention = Ruff 2 & 7 Selective Attention Test, Trial Making Test (Trials A/B) and Digit Span (from Wechsler Adult Intelligence Scale – Fourth Edition), Visual Recognition = Judgment of Line Orientation, Concept Formation = Wisconsin Card Sorting Task, 128 Card Version, Matrix Reasoning and Similarities (from the Wechsler Abbreviated Scale of Intelligence), and Verbal Memory = California Verbal Learning Test - Second Edition, Standard Form.

In an effort to determine which cognitive domains had the most variance, the mean variances of each measure were pooled into their respective cognitive domains (see Table 2) and then averaged. As can be seen in Figure 3, the construction and verbal memory domains demonstrated marked mean variance (>1.0 SD) across our high functioning sample, whereas the motor performance and visual recognition domains exhibited the least mean variability.
Figure 4.

Maximum Discrepancy and adjusted maximum discrepancy scores by participant for all neuropsychological measures.

Note. MD = Maximum discrepancy. The adjusted MD value does not take into account the highest or lowest performance score.

The Maximum Discrepancy (MD) between each individual’s highest and lowest \( z \)-scores was calculated using the methods described in Schretlen et al., (2003). In short, an individual’s highest \( z \)-score was subtracted from the lowest \( z \)-score to give the MD value. Thus, 33 \( z \)-scores were derived for each individual from the 10 neuropsychological tests employed in this study.

The MD values ranged from 2.26 to 7.06 (\( M = 4.14, SD = 1.25 \)) indicating that the lowest MD by any of the participants was 2.26 SDs and the largest was 7.06 SDs. Each participant’s MD values are illustrated in Figure 4. Figure 5 illustrates the frequency distribution of the MD values attained by our sample. As such, all of our sample produced MD values greater than 2.0 SDs or 30 standard score points, whereas, 64% (16 participants) obtained MD values greater than 3.0 SDs or 45 standard score points, 44% (11 participants) obtained MD values greater than 4.0 SDs or 60 standard score points, and 24% (6 participants) obtained MD values in excess of 5.0 SDs or 75 standard score points. Given that a few outliers may have inflated some of the MD values
within each participant, an adjusted MD score was calculated (see Schretlen et al., 2003), by removing the highest and lowest z-scores from each participant (Figure 4, dark bars). The adjusted MD values ranged from 1.79 to 5.14 ($M = 3.32$, $SD = 0.90$) indicating that the MD values were reduced with the lowest adjusted MD by any of the participants was 1.79 SDs and the largest was 5.14 SDs. None of the participants showed any MD value less than 1.0 SDs and as such did not express consistent performance (Schretlen et al., 2003).

Figure 5.

Frequency distribution of maximum discrepancy and adjusted maximum discrepancy scores.

In an effort to understand the relationship between IQ performance and performance on all the tests, difference scores were calculated between each measure and FSIQ-4 subtest $z$-score for each participant. The IQ $z$-scores ranged from 0.13 to 2.33 ($M = 1.49$, $SD = 0.69$) and the range of the average difference from IQ $z$-scores was from -2.58 to -0.34 ($M = -1.28$, $SD = 0.59$). Interestingly, all participants scored below their IQ performance on all measures. The IQ values showed a very strong negative correlation with the average difference score ($r = -0.75$, $p < 0.05$). As the scatterplot in Figure 6 illustrates, the average performance on other tests deviated negatively from the IQ $z$-score with increased IQ values. Additionally, the correlation between MD and the mean of each participant’s $z$-scores on all measures yielded a moderate negative value of -0.34 ($p < 0.05$). This indicated that as cognitive performance decreased, intraindividual
variability increased (Figure 7). These findings are consistent with Schretlen et al.’s (2003) results.

Figure 6.

Scatterplot depicting each participant’s FSIQ (in SD units) as a function of the mean of their mean difference from FSIQ on all test measures. The difference scores increased negatively with better FSIQ scores (Pearson’s $r = -0.75$).

\[\text{Note. FSIQ = Full Scale IQ – 4 subtest from the Wechsler Abbreviated Scale of Intelligence}\]
Figure 7.

Scatterplot depicting each participant’s maximum discrepancy (in SD units) as a function of the mean of their z-scores on all test measures. Intrindividuall variability decreased with better overall cognitive test performance (Pearson’s r = -0.35).
Chapter 4
Discussion

The purpose of this study was to determine (1) how much intraindividual variability would be evinced on a reasonably comprehensive battery of neuropsychological tests in a sample of healthy individuals with a particularly high level of educational achievement; (2) if the Best Performance Method would be an appropriate method to estimate pre-morbid intelligence in persons with particularly high level of educational achievement; (3) which (if any) cognitive domains would be most variable in persons with high levels of educational achievement; and (4) if healthy, highly intelligent persons evince test scores in the abnormal range.

In the simplest of terms, marked variability in neuropsychological test scores were found despite a rather uniform sample of highly functioning individuals. Discussion follows with respect to the individual research questions posed.

1 Intraindividual variability in a sample of healthy individuals with a particularly high level of educational achievement

With regards to interindividual differences, or diversity, the findings indicate that marked variability across all cognitive domains was observed in our high functioning sample. That is, there does not seem to be any consistency in test scores across cognitive domains. Furthermore, as shown in Figure 2, on nearly 50% of the neuropsychological measures used in the study, the mean variances were >1.0 SD. Variability was most evident on the GPT Trial 1 but it should be noted that previous research has illustrated little relationship between motor performance and intelligence (Lezak et al., 2004). Hence, it may be somewhat misleading to consider poor performance on this specific measure as equivalent to, say, poor performance on memory testing. The variability in performance on the DS and CVLT is consistent with previous research that illustrates that while low IQ individuals tend to have memory scores above expectation for IQ, the opposite is seemingly true of high IQ individuals in keeping with our study (Hawkins & Tulsky, 2001). Taken together, these finding are consistent with previous studies on normal, healthy individuals of average intelligence (Hawkins & Tulsky, 2003; Schretlen et al., 2003; Schretlen et al., 2008).
Marked quantitative discrepancies between neuropsychological scores have been argued to be normal, especially with highly intelligent populations (Dori & Chelune, 2004; The Psychological Corporation, 1997). For example, Tulsky and colleagues (2000) analyzed the base rates of WAIS-III FSIQ and index scores from the original normative sample of 2450 participants and concluded that as the level of FSIQ increased, the likelihood of large differences between index scores increased. They found that for those with FSIQ scores of ≥ 120, a 20-point difference between the Verbal Comprehension Index and the Processing Speed Index occurred in nearly 29%, whereas it occurred in just over 8% of those with FSIQ scores ≤ 79. Moreover, a 20-point difference between the Perceptual Organization Index and the Processing Speed Index occurred in just over 21% of the individuals with FSIQ’s ≥ 120 and just over 10% in individuals with FSIQ’s ≤ 79. These large discrepancies between index scores have continued with the new WAIS-IV (Wechsler, 2008). Future studies could focus on determining if this pattern continues with other neuropsychological test batteries.

The set of performance scores on the WASI (Vocabulary, Matrix Reasoning, Block Design and Similarities) along with their mean variances appear to be relatively consistent across all individuals as their mean z-scores all fall within one standard deviation of each other. It is important to note however, that the WASI is usually employed as a method to measure adult and adolescent intelligence and was originally designed to measure abilities whereas the other neuropsychological tests used in this study were designed to measure deficits (Schretlen et al., 2003; Wechsler, 1997). The WASI measures are also highly inter-correlated. Further research on the types of tests and what they actually measure is needed to clarify if such a relationship between tests measuring abilities versus deficits plays a significant role in the understanding of intraindividual variability.

2 The Best Performance Method in Persons with a Particularly High Level of Educational Achievement

With regards to the estimation of premorbid ability, our results indicated that the average person’s lowest z-score fell 2.58 SDs below their estimated IQ and that their highest z-score also fell below their estimated premorbid IQ, but by a factor of 0.34 SD. These findings suggest that a marked reduction of discrepancy scores is observed between an individual’s premorbid IQ and all of their cognitive test performances. These findings extend previous work by Schretlen et al.,
(2003) by further suggesting that, “downward” discrepancies between a person’s IQ estimate and their poorest cognitive test performance are more common than “upward” discrepancies. Our findings support the notion that normal, healthy individuals are not equally endowed with identical levels of ability across all domains of cognitive functioning (Hawkins & Tulsky, 2003; Schretlen et al., 2003, 2008). Of course, it is important to note that many (if not all) of these studies utilized a sample of normal, healthy, average intelligence adults whereas our study examined a uniquely high functioning sample.

Our results indicated that each person’s MD fell between 2.26 and 7.06 suggesting that our entire sample showed marked quantitative discrepancies greater than 1.5 SD. Excluding each person’s highest and lowest z-score on all 33 measures, the adjusted MD values continued to illustrate discrepancies between the second highest and second lowest measures of greater than 1.5 SDs. The Best Performance Method states that any test score that is ≥ 1.5 SD below the highest score is likely due to some abnormal pathology (Lezak et al., 2004). Since all individuals in our sample demonstrated at least one test score performance that was at least 1.5 SDs below the highest score, it would be in error if one were to employ the Best Performance Method to argue the presence of cognitive impairment (and hence deduce causality of cognitive impairment to an entirely unrelated event). Clearly, this would not be the case in our unique sample. Firstly, none of our participants reported a history of psychiatric, neurological or medical disorder (see exclusion criteria). Secondly, it has been demonstrated that the Best Performance Method frequently overestimates premorbid abilities (Binder et al., 2009; Reynolds, 1997), and thus, as suggested by Binder and colleagues (2009), Larrabee (2005), and Schretlen and colleagues (2003, 2008), abnormal performance on some proportion of the neuropsychological test scores are expected among stringent definitions of abnormality such as ≥ 2.0 SD, let alone ≥ 1.5 SD. Lastly, as noted by Crawford, Garthwaite, and Gault (2007), as the number of neuropsychological measures employed in an examination increase, the statistical probability of obtaining at least one low score, defined as a score ≥ 1 SD below the mean, increases. Similarly, Ingraham and Aiken (1996) theoretically reported that in a battery of only 6 neuropsychological measures, more than 20% of the sample would obtain abnormal scores when using ≥ 1 SD as the cut off. Based on these reasons, the current study concludes that the Best Performance Method is an invalid and unreliable method of estimating premorbid intelligence and should not be used in
a clinical setting particularly in patients who have obtained a high level of educational achievement.

It is worth noting also that it has been posited that the relationship between intelligence and the magnitude of intraindividual variability is correlated positively (Binder et al., 2009; Hawkins & Tulsky, 2003; Iverson et al., 2008; Matarazzo & Herman, 1985; Matarazzo & Prifitera, 1989; Schretlen et al., 2008; Tulsky et al., 2000; Wechsler, 1997, 2008) and negatively (Hultsch et al., 2002; Li et al., 2001; Li & Lindenberger, 1999; Rabbitt et al., 2001; Ram et al., 2005). Yet, these arguments have been made without due empirical investigation, or stated another way, without actually utilizing test score performance from persons who have achieved a high level of academic intelligence (i.e., persons with scores that are consistent with the extreme ends of the intelligence scale). The findings in this study contribute unique data to this debated relationship with suggestions that intraindividual variability decreases with better overall cognitive test performance (Figure 7). In other words, as intelligence increases, intraindividual variability decreases. It is important to note that the definition of intelligence in this instance refers to the overall score across all 33 measures. Certainly, future studies with control groups are needed to replicate and thus validate the current findings.

3 Variability Between Cognitive Domains in Persons with High Levels of Educational Achievement

The most variable cognitive domains measured from our high functioning sample were construction (RCFT and Block Design from the WASI) and verbal memory (CVLT) indicating that scores on these measures were detected to have the greatest range in test performance. A preliminary conclusion from our findings would suggest that, highly intelligent individuals are most prone to demonstrate variable scores on tests of construction and verbal memory. In contrast to our findings, a base rate study of the sample discrepancy scores from the WAIS-III and WMS-III normative sample performed by Dori and Chelune (2004) concluded that from the 1250 sample, large discrepancies between the verbal intellect and memory\(^2\) cognitive domains

\(^2\) For consistency purposes, the current study has adopted cognitive domain titles from Lezak et al., (2004). It is arguable, under various definitions of cognitive domains, that the CVLT may belong to the verbal memory and/or simply memory domains.
were observed. In their study, verbal intellect was shown to be superiorly high across all participants, whereas memory scores were substantially lower. Consistent with the findings of Tulsky et al. (2000), these discrepancies were even more pronounced in those individuals with more education. For example, a $\geq 15$-point discrepancy between Verbal IQ and General Memory was observed among 15% of those with 13-15 years of education. For those with 16+ years of education, this discrepancy increased to 21%. Although the current study’s findings stand in contrast to that of Dori and Chelune (2004), the relationship between cognitive domains and variability is still very much unknown and that all current data appears to be equivocal. Based on this and other equivocal findings on large discrepancy scores between measures, Sherman, Slick, Strauss, and Spreen (2006) have called for additional investigations of normal variability and its relation to cognitive domains.

Since no participants were included into our study that met criteria for a history of neurological or psychological disorder, yet marked impairment in construction and verbal memory cognitive domains was found, our results raise important concerns as it relates to the research literature that guides our clinical practice. For example, the results of previous studies have indicated that patients who have suffered from an mTBI (well after the expected time frame for recovery) performed significantly lower across all indices of the RCFT (a measure of construction) than the respective control group (Leininger, Gramling, Farrell, & Kreutzer, 1990; Raskin, Mateer, & Tweeten, 1998). The authors of these studies concluded that the cause of the difference between the test performances was likely due to the lingering effects of the mTBI. Clearly, our findings illustrate the importance of understanding intraindividual variability further if we are to indeed place diagnostic practice in clinical neuropsychology on firmer scientific grounds. Indeed further research replicating our findings and measuring other tests of construction and verbal memory are required so not to be exceptionally prone to false positive conclusions about cognitive functioning in neurobehavioral and neuropsychiatric disorders such as mTBI.

Interestingly, it is also worth noting that Levin (1995) has shown that the most impaired cognitive domains after a TBI are learning and memory and do not necessarily reach normal levels after 6 months of recovery. In contrast, the results of our study show that normal, healthy, highly intelligent persons may also exhibit “impairment” (i.e., impaired scores) in these domains. Our findings illustrate that where this may be particularly relevant, is within the clinical setting where a practitioner may be faced with large discrepancy scores between cognitive domains.
from a high functioning individual. In this instance, it may be erroneous to conclude that a pathological process is the cause of the impairment in these cognitive domains in healthy individuals as marked variability across test scores and indeed cognitive domains is not the exception, but the norm. This is especially true for those individuals with higher levels of education and intelligence as our findings demonstrate (see Figure 6). Thus, caution should always be exercised when attributing cerebral dysfunction as causal in the instance of several abnormal and largely discrepant scores without the consideration of historical, behavioural, diagnostic and current forensic (if applicable) situational information (Schretlen et al., 2003).

4 It is Possible to be Highly Intelligent Yet Neuropsychologically Impaired

The data from this study suggests that it is possible to be highly intelligent yet neuropsychologically impaired. To some extent, the results of this study are not surprising from both biological and environmental perspectives. That is, the equal endowment of cognitive abilities in any individual is highly unlikely, and even more unlikely is the consistency of this distribution of cognitive abilities across individuals. Thus, our study further illustrates that cognitive abilities are not equally distributed within individuals with high levels of educational achievement. Indeed, some participants in our sample achieved test scores at the very poor to borderline range coupled with scores in the above average to superior range on tests of neuropsychological ability.

As already noted, base rates of deficient neuropsychological test performance have been evaluated among cognitively healthy normal adults using a variety of measures (see Heaton et al., 1996, 2004; Palmer et al., 1998; Schretlen et al., 2003, 2008). Palmer and colleagues (1998) concluded that any strong assertions regarding the incidence of neuropsychological impairment appear difficult to justify solely on the basis of a few unrelated and “abnormal” test scores which is often the neuropsychological test pattern evinced by patients with mTBI, major depressive disorder, and schizophrenia (Zakzanis et al., 1999; also see Iverson, 2005). Moreover, it is also important to note that within the context of the neuropsychological evaluation of mTBI and neuropsychiatric disorders, the use of a large number of tests can also result in abnormal scores in entirely normal individuals (Binder et al., 2009; Larrabee, 2005). Therefore, an essential axiom in neuropsychology is the clinician’s reliance on findings that are not spurious or subject
to Type I error (Yedid, 2000). The absolute necessity of adherence to such strict clinical criteria is reflected in Heaton’s (1991, 2004) findings that 53% of non-neurologically impaired normal subjects performed in the dysfunctional range on 10% or more of the WAIS-R and Halstead-Reitan Neuropsychological Battery subtest scores. Furthermore, Ingraham and Aiken (1996) presented an empirical approach to determine abnormality in test batteries with multiple measures wherein it was found that within a battery with 30 scores in which the criterion is 3 tests falling at 1 SD below the mean, the probability was found to be nearly .90 that this would occur by chance alone. As Larrabee (2005) notes however, Ingraham and Aiken based their computations on independent scores, which is likely not the case in an actual test battery and as such, the probabilities are likely somewhat overestimated. It is clear that without the accountability provided by a base rate of dysfunction in terms of intraindividual variability, the scores of normal subjects on cognitive measures could potentially distort the clinician’s opinion. This would hold particularly true for a neurobehavioral disorder such as mTBI and neuropsychiatric disorders such as major depressive disorder and schizophrenia wherein base rates of cognitive impairments are low (Zakzanis et al., 1999). Our results suggest that even further caution is necessary in the instance of interpreting neuropsychological test scores produced by an individual with particularly high educational achievements.

In addition, our findings also contribute to the growing research on intraindividual variability. More specifically, our results support the body of research that suggests that as intelligence increases, intraindividual variability also decreases. In addition, the results of this investigation further support the notion that the level of intraindividual variability may also serve as a meaningful and systematic indicator of individual differences.

5 Limitations of the Present Study

Various limitations to this study are recognized. Firstly, the size of our sample was small. To this end, while a larger sample size may demonstrate homogeneity of cognitive test scores in terms of little to no variability, the alternative is just as probable in that a larger sample size may demonstrate further heterogeneity in terms of cognitive test scores. Additionally, given the occupational demand of our sample and resultant time constraints imposed by same, we also employed a rather limited number of neuropsychological tests. It is probable however that additional normal intraindividual variability would have been evinced on a more extensive
battery of neuropsychological tests in light of results garnered by Heaton et al., (1994, 2004), Palmer et al., (1998), Schretlen et al., (2003, 2008) and others in their respective samples. In addition, a further limitation perhaps inherent to the purpose of this study is that separate norms were used for each test and as such, variability across normative samples could have accentuated the appearance of within person variability. Accordingly, future research might employ a more comprehensive, integrated, modular battery of neuropsychological tests developed and normed together to assess a wide array of neuropsychological skills and functions with appropriate psychometric properties with extensive normative and validation data. Indeed, such batteries are now more widely available (e.g., the NAB, RBANS) and would perhaps serve to better measure change in function outside the bounds of normal variability. Lastly, future research should consider a more robust design. For example, it may be informative to study this high functioning population in a classic longitudinal test-retest paradigm to also ascertain the temporal stability (or variability) of test performance across time.

6 Conclusion

Despite these shortcomings, the findings of this study emphasize the importance of inferring patterns of cognitive functioning or impairment from sources other than psychometric variability. That is, performance in the context of other historical, behavior, or medical information may be just as, if not more important than psychometric data alone particularly when practitioners are often left without an appropriate heuristic when having to estimate a patient’s pre-injury cognitive functioning (Schretlen et al., 2003). Indeed, without readily available valid heuristics of pre-injury cognitive functioning such as pre-injury baseline testing, our findings suggest that we need to develop a better understanding of the relationship between various cognitive functions as articulated within our test data and real world abilities. Again, this is seemingly most concerning within the context of ascribing “cognitive impairment” to a neurobehavioral disorder such as mTBI, or neuropsychiatric disorders where base rates of cognitive impairment are low and educational achievements are high. Certainly, this task will often require a detailed analysis of specific activities in terms of daily living, and breaking them into their neuropsychological components. As Long and Collins (1997) suggest, in addition to gathering information about the cognitive components of everyday social and vocational activities, we need to further investigate and clarify the relationship between individual cognitive functions and specific test scores and patterns of results. Additionally, since the variability we found could just
as likely have been due to the differences in normative samples, these test scores should be compared to a homogenous normative sample. Clearly, if one were to blindly interpret our findings without gathering information about the cognitive components of everyday social and vocational activities engaged in by our sample, it is difficult to imagine that cognitive impairment would not be inferred from the marked variability of test scores across domains in this sample of, paradoxically, high functioning individuals with no subjective complaints of cognitive impairment.
References


as a guide for clinical and neuropsychological assessment. *Journal of Clinical Psychology* 45(6), 919-926.


Zakzanis, K.K. (2001). Statistics to tell the truth, the whole truth and nothing but the truth:
Formulae, illustrative numerical examples, and heuristic interpretation of effect size analyses for neuropsychological researchers. *Archives of Clinical Neuropsychology, 16*(7), 653-667.
