11-4-2024

Back to Baseline: Understanding the Complexities of Estimating Premorbid Cognitive Functioning in Sport-Related Concussion Assessment

Kassandra Helena Korcsog
University of Windsor

Back to Baseline: Understanding the Complexities of Estimating Premorbid Cognitive Functioning in Sport-Related Concussion Assessment

By

Kassandra H. Korcsog

A Dissertation
Submitted to the Faculty of Graduate Studies
through the Department of Psychology
in Partial Fulfillment of the Requirements for
the Degree of Doctor of Philosophy
at the University of Windsor

Windsor, Ontario, Canada

2024

Back to Baseline: Understanding the Complexities of Estimating Premorbid Cognitive Functioning in Sport-Related Concussion Assessment

by

Kassandra H. Korcsog

APPROVED BY:

____
D. Marshall, External Examiner
University of Michigan

____
K. Milne
Department of Kinesiology

____
J. Casey
Department of Psychology

____
R. Biss
Department of Psychology

____
C. Abeare, Advisor
Department of Psychology

October 7th, 2024

DECLARATION OF ORIGINALITY

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

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

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

ABSTRACT

In neuropsychology, a crucial aspect when working with people who have a known or suspected brain condition is the accurate estimation of their premorbid cognitive functioning. Within the context of post-concussion assessment for athletes, the primary objective is to determine whether and when a player has returned to their pre-injury cognitive baseline, thus judging their readiness to safely return to sport participation. Therefore, baseline testing is considered to be valuable and worthwhile for high contact sports with a high probability of sustaining a concussion. However, problems with baseline testing have also been identified due to unreliable integrity of baseline data. Hence, there is a growing recognition that the assessment of sport-related concussion could greatly benefit from advancements in the estimation of premorbid cognitive functioning. The objective of this dissertation was to refine the estimation of premorbid cognitive functioning by better understanding the demographic, personal, and educational variables that contribute to cognitive test performance.

To accomplish this, 158 athletes (M_{age} = 20.30, SD_{age} = 1.95, range = 17-25; 70% male; 64% White; M_{edu} = 13.20, SD_{edu} = 1.44) were recruited as a part of the University of Windsor's varsity athletics baseline testing effort from a range of men and women's sports (54% men's football). Each athlete completed a fixed battery of tests, beginning with consent and a questionnaire which asked for self-report of a broad range of demographic, personal, and educational information. Cognitive tests included in this study were the Immediate Post-Concussion Assessment and Cognitive Testing (ImPACT), the Test of Premorbid Functioning

(TOPF), the Trail Making Test A and B (TMT-A and TMT-B), and the Delis-Kaplan Executive Functioning System (D-KEFS) Verbal Fluency subtest (D-KEFS VF).

The first study acted as a general exploration of the variables that best related to cognitive test performance. Although several significant correlations were found, those that were between test scores and demographic, personal, and educational variables were generally weak. The second study aimed to understand the contribution of D-KEFS VF word quality (using word frequency values derived from the Corpus of Contemporary American English) produced during the letter fluency subtest, and it was determined that word frequency does not relate to test performance on the TOPF, ImPACT, or TMT-A and B, but personal variables such as GPA and race may play a role. The third study aimed to tie results from Study 1 and 2 together to generate prediction algorithms using standard simple and multiple regression analyses, but due to the weak relationships outlined in Study 1, regression analyses had poor predictive power across cognitive tests.

Overall, this dissertation has enriched the knowledge base of sport neuropsychologists by offering insights into the role played by demographic, educational, and personal variables in the interpretation and estimation of cognitive test performances. This research serves as a catalyst for advancing the sophistication and effectiveness of neuropsychological assessment methods and fosters a more personalized approach to evaluating cognitive functioning. It also informs that currently, approaches used to estimate premorbid cognitive functioning are not better than a baseline assessment in sport neuropsychology.

DEDICATION

This dissertation is dedicated to my parents, Jennifer and Béla Korcsog.

Your unwavering support, love, and encouragement have been my greatest inspiration. Such wholehearted parental involvement and guidance is a privilege that I am so fortunate to have. Thank you for believing in me and for being a constant source of strength.

This achievement is as much yours as it is mine.

ACKNOWLEDGEMENTS

There are several people that I'd like to thank for their contributions to this research project and to the successful completion of this degree. First, I would like to extend my heartfelt gratitude to Dr. Christopher Abeare for his transformative mentorship throughout my educational journey. Your support, insightful guidance, and profound expertise have not only shaped my research but have also influenced my personal and professional growth. You have consistently encouraged me to push boundaries and think critically, and your dedication to fostering an environment of positive intellectual curiosity has been invaluable. I am deeply appreciative of the countless hours that you have invested in my work and the confidence you have instilled in me. This achievement is a testament to your extraordinary mentorship, and I am so grateful to have learned so much from you.

To my exceptional committee members, Dr. Joseph Casey, Dr. Renée Biss, and Dr. Kevin Milne. Your expert guidance, thoughtful feedback, and strong support have been especially important for the successful completion of my PhD journey. Each of you has provided unique perspectives that have significantly enriched my research and have helped me to navigate complex challenges. To Dr. David Marshall, I am so appreciative of the time and effort you have dedicated to reviewing my work, offering your expert perspectives, and supporting my academic development during this final stage and as my external examiner.

To my family and friends, I want to extend my deepest gratitude to each of you for your support and encouragement throughout my many years of school. The countless moments of reassurance and understanding you have shown me have been a great source of strength and motivation. I am so grateful for your patience, sacrifices, and the joy you have all brought into my life, which has been instrumental in reaching this milestone.

Lastly, to Jake. Over the past 12 years your support as a boyfriend, fiancé, and now husband has been a foundational part of this journey. Your encouragement and sacrifices have meant the world to me, and I am so grateful for this life we're building together. Here's to many more years of shared dreams, achievements, and laughs.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

AACN – American Academy of Clinical Neuropsychology
ADHD – Attention-Deficit/Hyperactivity Disorder
AmNART – American Version of the National Adult Reading Test
ANT – Animal Naming Test
AoA – Age of Acquisition
CF – Category Fluency
CISG – Concussion in Sport Group
CS – Category Switching
DI – Demographic Information Estimation Formula Index
D-KEFS – Delis-Kaplan Executive Function System
EPGA – Estimated Premorbid General Ability
FSIQ – Full Scale Intelligence Quotient
GDP – Gross Domestic Product
GNT – Graded Naming Test
HART – Hopkins Adult Reading Test
ImPACT – Immediate Post-Concussion Assessment and Cognitive Testing
IQ – Intelligence Quotient
LF – Letter Fluency
LOFT – Lexical Orthographic Familiarity Test
mTBI – Mild Traumatic Brain Injury
NAART – North American Adult Reading Test
NART – National Adult Reading Test
NIH-TB-CB – National Institute of Health Toolbox Cognitive Battery
OPIE-3 – Oklahoma Premorbid Intelligence Estimate – 3
RIM – Rohling Interpretive Method
SES – Socioeconomic Status
SLD – Specific Learning Disorder
SLI – Specific Language Impairment
SRCC – Sport-Related Concussion Centre
TBI – Traumatic Brain Injury
TMT – Trail Making Test
TOPF – Test of Premorbid Functioning
VCI – Verbal Comprehension Index
VF – Verbal Fluency
WAIS-R/III/IV – Wechsler Adult Intelligence Scale – Revised/Third Edition/Fourth Edition
WCT – Word Choice Test
WISC-IV/V – Wechsler Intelligence Scale for Children – Fourth Edition/Fifth Edition
WMS-III/IV – Wechsler Memory Scale – Third Edition/Fourth Edition
WRAT-4/5 – Wide Range Achievement Test – Fourth Edition/Fifth Edition
WRAT-READ Wide Range Achievement Test – Reading Subtest
WTAR – Wechsler Test of Adult Reading

CHAPTER 1

Introduction

In the field of neuropsychology, many professionals follow the World Health Organization's *biopsychosocial framework for health* to guide their approaches (World Health Organization, 2001). The biopsychosocial model is interdisciplinary in nature, and considers biological, psychological, and social factors as well as their complex interactions in understanding health, illness, and healthcare in general. Biological factors include genetics and physiology, psychological factors include emotions, cognition, and behaviours, and social factors involve cultural, familial, and socioeconomic influences. Thus, it is an important theoretical framework to follow when providing a comprehensive understanding of patient care, as it integrates multiple dimensions of a person's life and health. Understandably, the biopsychosocial model is extremely relevant to neuropsychology because it provides a comprehensive framework for understanding the complex interactions that can affect brain function and behaviour.

Neuropsychology as a discipline is rooted in the study of brain lesions, which began in the late 19th and early 20th centuries. Early scientists such as Paul Broca (Broca, 1861) and Carl Wernicke (Wernicke, 1874) made significant strides by correlating specific brain injuries with particular cognitive and behavioural deficits, thereby mapping functions to brain regions. However, these early studies often relied on small, homogenous samples, predominantly composed of White, middle-to-upper class males, thus limiting generalizability of findings and introducing biases. Over time, the field recognized the need for more representative normative samples. Demographic corrections were introduced to account for variables such as age, sex, education, and race, leading to more accurate and equitable assessments of cognitive function across diverse populations (Kolb & Whishaw, 2003). This evolution reflects a broader trend in psychological research towards inclusivity and precision in understanding human cognitive and brain function.

Rivera-Mindt and colleagues (Rivera-Mindt et al., 2010) more recently challenged the discipline of neuropsychology to further improve normative sample development and increase access to competent neuropsychological services to all patients by diversifying the workforce and increasing cultural competence through multicultural training, education, and research. Given the increasing diversity of the population in North America and in-keeping with the biopsychosocial framework, the American Academy of Clinical Neuropsychology (AACN) has developed the *Relevance 2050 Initiative* (AACN, 2016) which aims to ensure that neuropsychological practices remain competent, relevant, and effective for all patients in the future. The Relevance 2050 initiative proposes that neuropsychology must support new assessment methods, training models, mid-career supervision models and clinical strategies to begin to substantially increase the percentage of patients that neuropsychologists can competently serve (AACN, 2016).

Following this, the AACN released a position statement on the use of race as a factor in neuropsychological test norming and performance prediction, outlining that the concept of *race* itself is often a proxy for factors that are attributable to inequity, injustice, bias, and discrimination (AACN, 2020). As such, they suggest that neuropsychology would benefit from an approach analogous to “precision medicine” where one’s expected cognitive performance is more accurately defined through the incorporation of a broader range of demographic and personal variables in prediction models, beyond more typical demographic inclusions such as race. Therefore, including a broader range of variables in prediction models related to methods of *estimation of premorbid cognitive functioning* could be vastly improved using these recommendations.

What is Premorbid Cognitive Functioning?

In neuropsychology, one of the most important considerations when working with a person with a known or suspected acquired brain injury or neurodegenerative condition is the estimation of their premorbid level of cognitive functioning (this is also referred to as premorbid intellectual functioning). The word *premorbid* is defined as "existing before the occurrence of physical disease or emotional illness" (Merriam-Webster Dictionary, 2021). Therefore, it is the normal or prior level of functioning against which a person's current, post-morbid state is compared. Clinical neuropsychological assessment often requires the comparison of a person's obtained test scores against some estimate of their premorbid level of functioning to determine the degree of decline that they have experienced, and this estimate can be derived in many ways. This estimate is referred to as the *comparison standard* and can be determined using normative data or individual characteristics depending on the person, the behaviour being evaluated, and the assessment's purpose (Lezak et al., 2012). Several approaches have been devised to estimate premorbid cognitive ability in patients, and the appropriateness of a given approach is likely to depend on the patient being investigated. As such, several comparison standards are discussed in detail.

Comparison Standards

The determination of deficits generally depends on the comparison between what is estimated to be the person's characteristic premorbid level of cognitive functioning and their current level of cognitive functioning. As such, this change is detected from historical data (including prior test scores if available), the obtained test performance scores, and qualitative features of the test performance (Lezak et al., 2012). This, in turn, is evaluated in the context of the presenting problem, recent history, the person's behaviour, and the neuropsychologist's knowledge of patterns of neuropsychological impairment (Lezak et al., 2012). There are several comparison standards that can be used to make this determination, including normative comparison standards, direct deficit measurement, and indirect deficit measurement.

Normative Comparison Standards

Lezak et al. (2012) discusses several different methods of normative comparison. First, they discuss the *population average*, stating that a score representing the average or median performance of a defined population is the normative comparison standard. Another is through *species-wide* performance expectations (also referred to as *criterion-referenced*, Sattler & Hoge, 2014), whereby many species-wide abilities arise early and similarly in all typically developing people (e.g., motor and visuomotor control), and do not typically change over time and with experience. Lezak et al. (2012) also discuss *customary standards*, which are normative standards that have been arbitrarily set (e.g., 20/20 vision).

Normative comparison standards can be helpful for many purposes in psychology, including as a description of one's cognitive status for educational and vocational planning (Lezak et al., 2012). However, when assessing a person with a known or suspected brain pathology, normative standards are appropriate only when the function or skill is being measured well within the capability of typical adults and does not vary greatly with age, sex, education, or general mental ability. Therefore, an example of an ability that would require an individual comparison standard would be vocabulary level, which correlates highly with both social class and education (Heaton et al., 2009; Rabbitt et al., 2007; Sattler, 2001). Thus, the first step required to measure one's level of cognitive deficit is to establish or estimate the person's premorbid performance level for all the functions and abilities being assessed (Lezak et al., 2012).

Direct Measurement of a Deficit

Direct deficit measurement is completed using individual comparison standards. This means that the examiner compares premorbid and post-morbid behaviour and evaluates the discrepancies, but this requires that premorbid test scores, school grades, or other relevant observational data is available (Lezak et al., 2012). In the sporting world, an example of this could be comparing an athlete's pre-season baseline neuropsychological assessment to their post-injury assessment. Prior neuropsychological assessment is typically assumed to offer the most reliable form of premorbid estimate of neuropsychological functioning (e.g., Langeluddecke & Lucas, 2004; Wilson et al., 1979). Unfortunately, this type of data is rarely available. However, in some populations previous neuropsychological testing may not always be the most accurate way to estimate premorbid functioning due to poor performance as a result of inadequate task engagement, poor motivation, lack of appreciation of the importance of doing their best, or distractions in the testing environment (e.g., baseline testing for athletes; Abeare et al., 2018; Tsuchima et al., 2019; Messa et al., 2020).

Indirect Measurement of Deficit

When completing indirect measurement of premorbid functioning, the examiner compares the person's current performance with an *estimate* of their original ability level, which can be inferred from various sources (Lezak et al., 2013). With that said, most techniques of indirect assessment of premorbid ability rely on cognitive test scores, extrapolation from current reading ability, demographic variables, or on a combination of these (Lezak et al., 2012). Lezak et al. (2012) describe several methods for estimating premorbid functioning including historical and observational data, mental ability scores, word reading tests, demographic variable formulas, and test score and demographic hybrid formulas.

Mental Ability Test Scores. A common method used to estimate premorbid ability level from test performance uses a vocabulary score as the single best indicator of premorbid functioning. This is because, according to the theory of fluid and crystallized general ability (Cattell, 1941; 1943), crystallized general ability loads highly on cognitive performances in which skilled judgement habits have become crystallized as a result of earlier learning application, and thus is maintained with age (Cattell, 1963). Fluid intelligence, on the other hand, is the ability to use logic and problem-solving in new situations without relying on pre-existing knowledge and thus is subject to environment, particularly brain damage and age (Cattell, 1971). Therefore, this method was derived from the observation that many individuals who were cognitively declining due to illnesses such as dementia retained well-established verbal skills long after recent memory, reasoning, arithmetic ability, and other cognitive functions were compromised. These tests, in addition to word reading tests, are helpful because they are brief estimates of premorbid intelligence that are relatively immune to neurological insult (Green et al., 2008; McGurn et al., 2004), thus providing an estimate of the "starting point" for evaluating cognitive changes. This is why they are commonly referred to as being "hold tests," as they are relatively unaffected by most forms of neuropathological change and can "hold" an individual's level of functioning (Russell, 1980).

The mental ability score method estimates premorbid functioning by using performance on tests that are highly correlated to IQ and resistant to effects of brain damage (Lezak et al., 2012). For the Wechsler Adult Intelligence Scale – Fourth Edition (WAIS-IV; Wechsler, 2008), the Vocabulary subtest score is generally identified as the best estimate of premorbid IQ because it is the subtest most highly correlated with the Verbal Comprehension Index (VCI) and Full-Scale IQ (FSIQ), as well as with level of education. Some other examples of tasks used for this purpose are the Shipley Institute of Living Scale (Shipley, 1940), Deterioration Ratios (comparing scores on Vocabulary from the WAIS-IV and other verbally weighted scores with performance on tests sensitive to attentional deficits and visuomotor slowing), and the combination of both Vocabulary and Picture Completion on the WAIS-IV (Wechsler, 2008). A major weakness of the mental ability score method is the false assumption that these subtests are resilient to most brain injuries. For example, conditions such as aphasia or dementia often affect the expressive language skills required for the Vocabulary subtest of the WAIS-IV (Lezak et al., 2012).

Word Reading Tests. Word reading tests are used because they are considered to provide an estimate of vocabulary size. These tasks also inform expected performance on tests of fluid intelligence like memory, executive functions, and processing speed, though their predictive value varies across cognitive domains (Duff et al., 2011; Duff et al., 2019; Schretlen et al., 2005). It is notable that some injuries may occur throughout the lifespan, including in youth, adolescence, and young adulthood (Faul & Coronado, 2015), and could alter trajectory or stunt acquisition of skills like word-reading and phonetic processing. The impact of repetitive head injuries is also an emerging concern, which is suspected to have similar lifespan influence, particularly for those who play sports, certain military personnel, and physical abuse survivors (Maroon et al., 2015).

The National Adult Reading Test (NART) was designed to estimate premorbid abilities of older adult patients in Great Britain suspected of suffering from dementia, and it is based on word recognition of irregular English words (Nelson, 1982; Nelson & Willison, 1991). It has been shown, however, that the NART's correlations with other cognitive domains such as executive function, memory, visuospatial, and perceptual-motor functions are significantly lower than with IQ scores, limiting its usefulness with respect to these domains (Lezak et al., 2012). The North American Adult Reading Test (NAART; Blair & Spreen, 1989) is a modification of the NART which was developed to be used for a North American English-speaking population and consists of 61 irregular and infrequently used words for comparison against the WAIS-R (Wechsler, 1981). Additionally, the American National Adult Reading Test (AmNART) is a 50-word version of the NART that was developed to be appropriate for the ethnically heterogeneous US population (Storandt et al.,1995; Strauss et al., 2006).

Two measures that are frequently used clinically to assess premorbid intellectual ability more recently include the Wide Range Achievement Test Reading Subtest (WRAT-READ; Wilkinson, 1993; Wilkinson & Robertson, 2006; Wikinson & Robertson, 2017) and the Test of Premorbid Functioning (TOPF; Pearson Clinical, 2017) which was previously known as the Wechsler Test of Adult Reading (WTAR; Psychological Corporation, 2001) (Berg et al., 2016). The Wide Range Achievement Test, 4th Edition (WRAT-4; Wilkinson & Robertson, 2006) is a standardized academic achievement battery consisting of four subtests. While designed specifically as a test of reading achievement, the word reading subtest is also one of the most selected reading tests used to estimate premorbid intellectual ability. The word reading subtest of the WRAT-4 consists of 55 words that are a mixture of standard and irregular words that are presented in order of descending frequency. Validation studies indicate that the WRAT-4 READ is highly correlated with the Wechsler Adult Intelligence Scale, Third Edition (WAIS-III; Wechsler, 1997) FSIQ (r = .71), and has been effective in estimating premorbid abilities for patients with traumatic brain injury (Johnstone & Wilhelm, 1996), drug abuse (Ollo et al., 1995), schizophrenia (Weickert et al., 2000), and Huntington's Disease (O'Rourke et al., 2011). WRAT-4 READ also provides a better estimate of the lower ranges of verbal intelligence, making the WRAT more applicable to the population at higher risk for traumatic brain injury (Wiens, Bryan, & Crossen, 1993). The WRAT-4 was revised to its newest edition, the WRAT-5, in 2017 (Wilkinson & Robertson, 2017), and the word reading subtest is still frequently used as a test of premorbid functioning by many clinicians. The only difference between the WRAT-4 and WRAT-5 regarding word reading is that the publishers added lowercase letters as the first items before the words, which represent letter sounds as correct responses for younger ages. This precedes word reading and is not administered to adults unless their raw score on word reading is less than five, corresponding to extremely poor reading ability.

The Wechsler Test of Adult Reading (WTAR; Psychological Corporation, 2001) is a list of 50 phonetically irregular words used to estimate premorbid intellectual functioning, using the same normative sample as the WAIS-III and WMS-III (Wechsler Memory Scale – Third Edition). The Test of Premorbid Functioning (TOPF; Pearson Clinical, 2017) is a revised and updated version of the WTAR that also relies on phonetically irregular words that are presented with descending frequency, therefore increasing difficulty. It uses a combined demographics and reading prediction equation involving both performance on the word-reading task and several demographic variables to derive an estimated premorbid FSIQ. The WTAR has been the subject of a considerable amount of research showing that the WTAR estimated FSIQ scores were lower than expected, particularly for individuals with dementia and brain injury (Leritz et al., 2008; McFarlane et al., 2006). Thus, a revision goal for the TOPF was to obtain a more accurate estimate of premorbid ability by reducing the effect of dementia and brain injury on combined demographic and word-reading equations (see Pearson Assessment, 2009). To improve the prediction range, the revision of the TOPF included increasing the number of words in the test and adding words with higher difficulty and adding occupation level to the existing set of demographic variables, which had already included region, sex, race/Ethnicity, education, personal factors, and developmental factors. One study by Shura et al. (2020) demonstrated that the TOPF is moderately to highly correlated with the WAIS-IV FSIQ: r = .56-.73 (Wechsler, 2009). Estimates of premorbid intelligence obtained from the TOPF and the WRAT-READ also have a strong linear relationship but have been shown to generate inconsistent estimates in a neurodegenerative disease clinical sample, and should not be used interchangeably (Berg et al., 2016). In general, single word reading tests tend to perform best for estimating verbal IQ and Full-Scale IQ and for scores in the average range, whereas they are less accurate for estimating performance IQ, IQs at lower and superior levels, and for people with learning disabilities or dementia (Lezak et al., 2012).

Prospective studies of word-reading ability after moderate-to-severe TBI demonstrate either stable performance or acutely lower scores compared to controls that then improve over time (Green et al., 2008; Mathias et al., 2007b; Steward et al., 2018). However, mild TBI (mTBI) populations show no difference in word-reading 1-month post-injury compared to controls despite lower scores on memory, processing speed, and set-shifting tests, supporting the validity of word reading as a hold test in mTBI populations (Steward et al., 2018). In one study by Joseph et al. (2021), the authors found that the TOPF frequently underestimated post-injury intelligence in a traumatic brain injury population, and the authors concluded that it is therefore not accurately measuring premorbid intelligence in their sample, particularly in those with above average to superior intelligence.

Other Word-Based Tests for Estimating Premorbid Ability. There are cases in which ability for oral reading is limited, such as in elderly persons with stroke or dementia. As such, some examiners use word recognition tests to aid in the assessment of premorbid ability. Spot-the-word is one of the two tests in *The Speed and Capacity of Language Processing Test* (McFarlane et al., 2006) developed to evaluate cognitive slowing following brain damage. The task is to identify the real word in each of 60 pairings of word and nonword. The *Lexical Orthographic Familiarity Test* (LOFT; Leritz et al., 2008) also uses a paired forced-choice format, but the choice is between words on the WTAR list and the same length unfamiliar English words. People tend to score higher on the LOFT than the WTAR, and it is often recommended for language-impaired persons. One advantage of using a word recognition reading strategy is that word recognition reading is more resistant to deterioration from brain damage and correlates highly with level of education (Lezak et al., 2012).

Verbal Fluency. Tests of verbal fluency are often a component of neuropsychological test batteries, especially in cases involving damage to the left frontal lobe, which is known to be associated with phonemic fluency, or to the temporal structures, which are known to be involved with semantic fluency (Henry & Crawford, 2004). Naming ability also relies upon input from temporal-lobe based semantic networks and frontally based phonemic retrieval processes (Melrose et al. 2009). However, the influence of age (e.g., Acevedo et al., 2000; Barry et al., 2008) and education (Dursun et al., 2002; Moraes et al., 2013) on verbal fluency performance have been described to be pertinent to performance. Several studies have reported significant correlations between verbal fluency and premorbid intellectual ability as measured by oral word reading. Crawford et al. (1992) found a highly significant correlation of .67 between phonemic fluency and the national adult reading test (NART; Nelson, 1982). Harnett et al. (2004) reported a correlation of .47 between phonemic fluency and NART and a correlation of .33 between semantic fluency and NART. Ardila et al. (2000) suggest that these correlations are unsurprising since tests of verbal fluency and oral word reading are both dependent on verbal ability, and both tests are closely linked with verbal intelligence.

A study by Jenkinson et al. (2017) aimed to bring attention to the need for developing predictive methods for cognitive abilities other than general intelligence and focused on developing regression equations for the prediction of verbal fluency and naming ability using the Test of Premorbid Functioning (TOPF; Wechsler, 2009) as the predictor. They developed a regression equation using the TOPF in order to establish how well the TOPF score and typical demographics (age and education level) predict performance on three commonly used verbal neuropsychological tests, FAS (a form of Controlled Word Association Test; COWAT; Benton et al. 1994), the Animal Naming Test (ANT; a test of semantic fluency; Goodglass & Kaplan, 1983), and the Graded Naming Test (a measure of naming ability; McKenna & Warrington, 1980; GNT). They found that the regression equations they developed were similar in terms of the amount of variance they accounted for, to other equations that have been reported calculating premorbid abilities from other neuropsychological tests (e.g., Harnett et al., 2004; Knight et al., 2006). Therefore, the findings of this study provide further evidence of the utility and relative efficiency of using regression equations for predicting scores on tests other than IQ tests using the TOPF but could also demonstrate the utility of verbal fluency as a measure of premorbid functioning.

No known published research has demonstrated the utility of verbal fluency as a direct measure of premorbid functioning. A poster presentation by Abeare and Seguin (2014) aimed to determine whether information about the quality of participant responses on a phonemic fluency test (FAS) was related to test-taker intelligence. To do this, FAS verbal fluency test responses and North American Adult Reading Test (NAART) scores of 85 participants were collected, and a number of Age of Acquisition (AoA; i.e., the age at which a word is typically learned) indices were generated based on a normative database of AoA values (Cortese & Khanna, 2008; Kuperman et al., 2012) in order to determine if these measures were related to estimated IQ based on NAART performance. They found that those with a higher NAART estimated IQ were more likely to generate words on FAS that had a higher AoA value, that maximum AoA produced is a better predictor of estimated IQ than the NAART itself, and that the AoA index provides unique information about test-takers. As such, the AoA index or other related word quality information might serve as an important predictor variable when developing an estimate of one's premorbid functioning.

Demographic Variables for Estimating Premorbid Ability. To make conclusions about the effects of a neurological condition on cognitive functioning, one must consider factors known to be associated with normal cognitive test performance variability. It has been well-established that several different demographic variables have a strong relationship with IQ (Matarazzo, 1972), including age, sex, race, education, and occupation. Proxy measures of cognitive reserve such as socioeconomic status are often considered as well (Heaton et al., 1991). However, these factors are much more complex than they might first appear. For example, measuring educational quality is much less straightforward than using the standard number of years of education and potentially encompasses interrelated factors like achievement, school-specific features (e.g., teacher-student ratio, teacher experience/salary), and regional SES. No known studies have included the more specific and complex factors such as these when seeking to improve the estimation of premorbid function using demographic variables.

As outlined by The American Academy of Clinical Neuropsychology (AACN)'s Position Statement on the use of race as a factor in neuropsychological test norming and performance prediction (AACN, 2021), race as a concept has a problematic history within Western science, as the term *race* lacks a clear definition. The authors describe that race is best viewed as a social construct that maintains a particular sociopolitical hierarchy and that historically has justified or excused cruelty, discrimination, exclusion, and exploitation. As such, genetic science has shown that greater variability exists within than between Racial groups, thus rendering the construct less meaningful for human biology. The AACN recommends that neuropsychology would benefit from an approach analogous to "precision medicine" in which clinicians should attempt to define expected performance more accurately through incorporation of a broader range of variables in our prediction models.

Demographic Variables Combined with Test Scores for Estimating Premorbid Ability. Some strategies have used both demographic variables and test data to predict premorbid IQ via a regression formula. Barona et al. (1984) were some of the first to attempt to standardize demographic prediction of premorbid IQ by developing a regression formula based on variables of age, gender, race, occupation, education, urban versus rural settings, and region (demographic information estimation formula index). The Barona method has demonstrated good prediction for average range scores (Lezak et al., 2012). The Oklahoma Premorbid Intelligence Estimate – 3 (OPIE-3; Schoenberg et al., 2006), Hopkins Adult Reading Test (HART; Schretlen et al., 2009), and Wechsler

Test of Adult Reading (WTAR; Holdnack, 2001) are other examples of this strategy. The OPIE procedure combines both premorbid demographic variables of age, education, occupation, and race with current performance on the WAIS-R Vocabulary and Picture Completion subtests in estimating premorbid IQ, the HART oral word reading performance was combined with demographic variables to generate regression equations that predict IQ scores obtained concurrently and 4–8 years earlier, and the WTAR, similar to the HART, combines word reading performance and demographic variables as well to predict IQ.

Notably, regression to the mean affects all these methods. For example, the demographic information estimation formula index (DI) developed by Barona et al. (1984) has consistently shown to provide a quick and accurate estimation of premorbid abilities for most subtests, but it tends to underestimate individual IQ's who are above 125 and over-estimate individual IQs that are below 75 (Barona et al., 1984; Reynolds, 1997). However, this restriction in range would be expected for any regression-based method, as these normally lead to less variation in the estimated scores (Helmes, 1996). A study by Powell et al. (2003) found that the DI appeared to provide the most clinical utility as an estimate of premorbid intelligence in a cognitively impaired sample in comparison to the OPIE (Krull et al., 1995).

Miller and Rohling (2001) presented a model of data interpretation that used a measure of premorbid functioning that they referred to as the Rohling Interpretive Method (RIM). This method uses a general measure of premorbid functioning termed the Estimated Premorbid General Ability (EPGA), which is an average of available premorbid measures, including school records, class standing, achievement test scores, preinjury testing, NAART, TOPF, Barona estimate, and any other premorbid estimates that might be available. The RIM uses objective statistical methods (i.e., t-tests, a test of heterogeneity, confidence intervals) to compare individual performances on neuropsychological domains of function to their EPGA. Notably, the EPGA was not intended to be a predictor of FSIQ but was designed to represent an individual's comprehensive premorbid general ability on neuropsychological tests encompassing a variety of cognitive skills, including memory, executive functioning, and motor skills (Miller & Rohling, 2001). Overall, the RIM allows for a quantitatively based comparison of an overall battery of measures and creates a greater reliance on quantitative methods of data interpretation. However, several critiques have been made regarding the RIM as outlined by Palmer et al. 2004. These include: a failure to distinguish statistically significant from pathological differences, an assumption that declines in specific abilities can be inferred when a particular test score deviates from an estimate of general premorbid ability, and confusion between the standard deviation associated with individual test scores versus that of a composite of those scores.

Crawford et al. (2001) conducted a study whereby they assessed the accuracy with which clinicians estimate premorbid IQ from demographic variables and compared it with a regression equation which uses the same information. Sixty participants were administered the WAIS-R and had their demographic variables recorded (age, sex, years of education and occupation). Eight clinical psychologists estimated the participants' IQs from the demographic variables, and the estimated IQs were also obtained using a regression equation developed by Crawford and Allan (1997). They came to find that demographic-based regression equations can provide unbiased and useful estimates of premorbid IQ, and these estimates can be modified in the light of additional qualitative information available to the clinician.

The Best Performance Method. Lezak et al. (2012) recommend the best performance method for estimating premorbid IQ, which bases premorbid IQ on the strongest indicator of intelligence across different categories of behaviors. Some potential sources of data include current or past test scores, relevant behavioural observations (i.e., verbal language skills), or evidence of premorbid achievement (i.e., school grades, occupation, family reports, army rating, or other types of awards that could suggest special skills or intellect). One major advantage of this method is that clinicians can use a breadth of information on which to base their IQ estimate, as well as the flexibility it affords in decision making. A major disadvantage is the clinician bias, often towards overestimating premorbid ability (Crawford et al., 2001).

To reduce the potential for overestimation, Lezak et al. (2012) warned that clinicians should not rely on a single test score to estimate IQ. The best method therefore rests on several assumptions that should guide the clinician in its application: First, that there is one performance level that best represents each person's cognitive abilities and skills generally. Second, that marked discrepancies between the levels at which a person performs different cognitive functions or skills give evidence of a condition that has interfered with the full expression of the person's cognitive potential. Third, that cognitive potential or capacity of adults can be either realized or reduced by external influences. Fourth, that few people consistently function at their maximum potential. Fifth, that within the limits of chance variations, the ability to perform a task is at least as high as a person's highest level of performance of that task. Lastly, that a person's premorbid ability level can be reconstructed or estimated from many kinds of behavioral observations or historical facts.

Implications of The Flynn Effect

An important consideration to be made when estimating premorbid functioning is that of the Flynn Effect. The Flynn Effect is a phenomenon of increased IQ estimates in more than 30 countries that developed over several decades (Flynn, 2007). Flynn (2013) attributed this to factors associated with modernization, including increased formal educational opportunities for preschoolers, higher numbers of college-educated adults, urbanization, development of a visual culture, more creative work roles, more leisure, better nutrition, and smaller family sizes. These increases in IQ are typically seen in fluid intelligence associated with differential patterns of improvements depending on a country's industrialization history. For countries that modernized before the 20th century, IQ gains seem to average about 3 points per decade, whereas countries modernizing in the mid- or later 20th century have demonstrated more significant gains (e.g., Kenya, 8 points).

In contrast, some countries that began modernization in the 19th century have reached an asymptote, with negative gains during the past 2 decades (e.g., Norway, Britain, and Sweden). An implication is that some developing nations may close the gap by 2050, while countries that do not modernize will continue to lag behind (Flynn, 2013). Therefore, premorbid estimates produced by tests of premorbid abilities that were designed to predict earlier editions of the Wechsler scales (such as the WTAR) will tend to inflate the level of premorbid baseline, leading to the possibility of a spurious diagnosis of deficit (Norton et al., 2016). Therefore, clinicians should always employ the most recently normed versions of these tests, using the most relevant norms to the country their client was educated in.

Canadian Versus American Normative Comparisons

Another important consideration to be made is of which normative data to use to score neuropsychological tests. Oftentimes Canadian psychologists must choose whether to combine scores from different normative samples when assessing a client or converting all the raw scores in an assessment to a common metric (i.e., American normative data), which also means that they avail themselves to a more extensive body of research regarding known score patterns for specific disorders. For example, the American normative data for the Wechsler Adult Intelligence Scale – Fourth Edition (WAIS-IV; Wechsler, 2008) are based on a much larger sample, it is co-normed with other test batteries (such as the Wechsler Memory Scale – Fourth Edition and Wechsler Individual Achievement Test – Second Edition), it includes both age and demographically adjusted normative scores, and numerous studies (e.g., Chaudhry & Ready, 2012; Heyanka et al., 2013) and book chapters (e.g., Brooks et al., 2013; Cullum & Lacritz, 2009) are published that use these normative data.

This issue has been a challenge for over 25 years since Canadian psychologists have questioned the appropriateness of American normative data for use in Canada (e.g., Beal, 1988). This prompted the Pearson Corporation to evaluate whether data from two countries were equivalent, which resulted in the test publisher developing separate Canadian normative data (Wechsler, 1996, 2001, 2003, 2004). What was found was that on the WISC-III and the WISC-IV, Canadian children obtained higher raw scores than American children (Beal et al., 1996; Wechsler, 2004). Similarly, on the WAIS-III and the WAIS-IV, researchers reported that Canadian adults obtain higher raw scores than American adults, although these differences are less apparent in adults over the age of 65 (Wechsler, 2008). The developers of the Wechsler scales hypothesized that these differences were due in part to variations in population composition between the two countries with respect to race, Ethnicity, and educational attainment (WAIS-IV Canadian technical manual, Wechsler, 2008).

A study by Harrison et al. (2014) compared the interpretive effects of applying American versus Canadian normative systems in a sample of 432 Canadian postsecondary-level students who were administered the WAIS-IV as part of an evaluation for a learning disability, attention-deficit hyperactivity disorder, or other mental health problems. Employing the Canadian normative system yielded IQ, Index, and subtest scores that were systematically lower than those obtained using the American norms. Notably, the percentage agreement in normative classifications, defined as American and Canadian Index scores within five points or within the same classification range, was between 49% and 76%. Therefore, *substantial* differences are present between the American and Canadian WAIS-IV norms, and the authors of this paper urge clinicians to carefully consider the implications regarding which normative system is most appropriate for specific types of evaluations. This is especially relevant in varsity athletics, given that athletes come from different provinces or countries to play their sport for a university or college.

Neuropsychological Assessment of Sport-Related Concussion

Sport-related concussions are a salient public health concern, with 1.1 – 1.9 million sport- and recreation-related concussions occurring annually in children 18 years of age or younger in the United States alone (Bryan, et al., 2016), and there being a risk of concussion in nearly every sport (Clay et al., 2014). In Canada, 1 in 450 people over the age of 12 were found to report sport-related concussions and other brain injuries occurring while engaged in sports or physical exercise as their most significant injury associated with disability (Gordon & Kuhle, 2022). Dr. Tator from the Canadian Concussion Centre estimates that the number of concussions sustained by Canadians each year is around 200,000, and according to Dr. Schneider from the University of Calgary, this number may be closer to 250,000 (Government of Canada, 2019). However, these estimates include only the number of people who have presented themselves to emergency departments, and so are likely an underestimate when considering those that do not seek medical attention.

Concussions are complex injuries that are difficult to diagnose and manage, especially in sport settings where recovery is largely dependent upon the interventions pursued at the recommendation of the athletic team (Kroshus et al., 2015). Evidence of poorer cognitive health among retired athletes with a history of concussion is evolving, and Cunningham et al. (2020) found that a history of sport-related concussion may more greatly affect memory, executive function, and psychomotor function.

Neuropsychological testing is commonly used as a way of detecting the neurocognitive effects of sport-related concussion and provides insight into whether an athlete should be returning to their sport. The medical management of sport-related concussion can be conceptualized as having two distinct components: the acute care management of the injured athlete (i.e., identifying and treating any neurosurgical emergencies such as a hemorrhage), and the monitoring of symptoms over time for tracking recovery and making return-to-play decisions. Many sport-related concussions, however, produce several subjective symptoms (e.g., headaches, dizziness, changes in balance/coordination, memory impairment) that could last for days or weeks post-injury (Erlanger et al., 2003), and most return-to-play guidelines agree that athletes should be symptom-free before returning to play (McCrory et al., 2017).

Although common cognitive deficits following a concussion are most often found in the areas of attention, processing speed, executive functioning, and memory (Rabinowitz & Levin, 2014), given the diffuse nature of concussion, other deficits can result as well. Karr et al. (2014) conducted a systematic review of meta-analyses which showed "staggering variability" in effect sizes across studies, thus speaking to the heterogeneity of concussion sequelae, and the importance of individualized assessment of the person who has sustained a concussion. Many studies have also been published on the time course of recovery from concussion, and predictors of prolonged recovery. McCrea et al. (2012) found that 10% of athletes experienced a protracted symptom recovery (i.e., greater than seven days) that was associated with longer recovery on neurocognitive testing. Eisenberg et al. (2013) found that in patients presenting to the emergency department with concussion, those who had a history of a prior concussion experienced symptoms for a longer duration (24 days) than those with no prior concussion history (12 days).

Demographic and symptom variables that predict the course of recovery from concussion have also been studied. McCrea et al. (2012) concluded that ultimately, there is significant heterogeneity in the literature regarding both the expected time course of recovery from concussion, as well as demographic and other factors that predict the course of recovery. Generally, it appears that younger athletes require more recovery time than older athletes (Field et al., 2003), and greater symptom severity at the time of injury predicts longer recovery (Meehan et al., 2013). Given this variability, an individualized approach in concussion management is warranted, just as Karr et al. (2014) recommend.

Temple (2019) noted that typically the neuropsychological evaluation for a sport-related concussion begins with a targeted clinical interview. Factors that suggest the relative severity of the concussion are obtained, including whether or not the individual lost consciousness, and the duration of retrograde and posttraumatic amnesia (the former meaning the inability to recall information that occurred from the point of the concussion backward, and the latter the ability to recall information from the concussion forward in time). In addition to injury-specific information, it is important to collect information about pre-existing learning disorders, Attention-Deficit/Hyperactivity Disorder (ADHD), psychiatric disorders, and other conditions that would help estimate the athlete's level of premorbid functioning from which to compare results that are obtained.

It is also important to consider the developmental stage of varsity athletes, who are typically between the ages of 18-25 and whose brains are still developing. This time is often referred to as "emerging adulthood" and describes a distinct period of development characterized by exploration and transition (Arnett, 2000). As such, when completing any kind of neuropsychological assessment with an athlete of this age, it is important to consider the current and continuing development of cognition, emotional regulation, and identity formation, as well as challenges with uncertainty and anxiety, economic pressures, and social and cultural expectations (Arnett, 2000).

Baseline Cognitive Testing

Neuropsychological assessment can occur at different points in the pre- and post-injury process. As mentioned previously, normative information is typically used to determine one's level of performance compared to some expectation, due to the lack of information about the person's true premorbid level of functioning. Temple (2019) discusses that in the case of athletes in high contact sports with a high probability of sustaining a concussion, obtaining a measure of baseline performance is considered to be valuable and worthwhile. Generally, comparing one's own baseline to a post-concussion performance is superior to comparison to a normative sample, because this allows for the direct comparison of their premorbid abilities rather than using an estimation. Baseline assessments can be especially helpful in the case of individuals with premorbid conditions that adversely affect cognitive test performance, such as ADHD or a learning disorder.

Baseline testing most often involves the computerized administration of a relatively brief (30 minutes or less) battery of tests measuring cognitive abilities known to be sensitive to the effects of concussion (e.g., attention, processing speed, memory). However, one of the greatest threats to the validity of a neuropsychological test is the effort or motivation of the test-taker. In the case of an athlete at baseline, it is plausible that they may be motivated to produce a very low baseline test score, so that they can meet or exceed that score when they are re-tested following a suspected concussion (this is known as "sandbagging"). For this reason, neuropsychological testing should always include validity checks to ensure that the test-taker is performing to the best of their ability.

Problems with Baseline Testing. Return to play guidelines strongly recommend that athletes not return to play until they are completely asymptomatic and fully recovered cognitively (Broglio et al., 2014; McCrory et al., 2017), and baseline testing often helps to operationalize the meaning of "fully recovered" for each individual athlete (Piland et al., 2010). However, Erdal (2012) points out that the utility of the comparison between post-injury and baseline test data in return-to-play decisions is based upon the integrity of the baseline data. That being said, baseline data is not always ideal, and Iverson and Schatz (2014) conclude that there is insufficient evidence that having baseline test results to compare to is superior to not having baseline test results in the sporting world.

The Immediate Post-Concussion Assessment and Cognitive Testing (ImPACT; Lovell, 2022), one of the most widely used computerized neurocognitive tests at baseline, includes several built-in indicators of potentially invalid baseline data. A systematic review by Gaudet and Weyandt (2016) found invalidity rates (based on ImPACT invalidity indicators) in normal baseline samples ranging from 2.7% to 27.9% with a weighted prevalence rate across the 12 studies of 6.1%. An extension of this review by Messa et al. (2020) found that when looking at prevalence of invalid performance at baseline testing using ImPACT, 6% of baseline assessments were found to be invalid by the ImPACT's default EVI, and between 22-35% were flagged as invalid by alternative EVIs. Therefore, the base rate of invalid performance from athletes at baseline is well above 6% and suggests that alternative validity indicators be used for these assessments. A study by Abeare et al. (2018) evaluated base rates of failure on published validity indicators on the ImPACT, which were compared within and across age groups. They concluded that most of the sample (55.7%) failed at least one of four validity indicators.

Schatz's (2010) study of collegiate athletes' ImPACT baselines (taken at a 2-year interval) found 2-year test-retest reliability to be higher and in the good range for speed composites (Processing Speed ICC – 0.74; Reaction Time ICC = .68) and the Visual Memory Composite (ICC = .65), than for Verbal Memory composite (ICC = .46). Other studies have questioned the acceptability of this test-retest reliability given that .70 is the normally applied cutoff for acceptable reliability (Mayers & Redick, 2012). ImPACT also does not come with standardized instructions to be given to athletes before test administration, which may contribute to variability and increased error (Moser et al., 2011). Additionally, ImPACT can be administered in a group or individualized setting, and while group administration can be more appealing due to fewer time and personnel requirements, it can also be detrimental to the validity of test data (Moser et al., 2011).

An alternative test to consider for baseline testing is the NIH Toolbox Cognition Battery (NIHTB-CB; NINDS 2017), a computerized neuropsychological screening battery that can be used as a brief, efficient, and reliable cognitive assessment in clinical practice and research. It measures the following cognitive functions: executive function, attention, episodic memory, language, processing speed, and working memory. The seminal article on the Toolbox was written by the developers in 2013 (Gershon et al., 2013), and since then numerous studies have been published on its use as a screening assessment. Abeare et al. (2021) developed embedded validity indicators for the NIHTB-CB using the Medical Symptom Validity Test (MSVT) as criterion. The EVIs developed effectively discriminated between patients who passed versus failed the MSVT, and the authors found that aggregating EVIs within the same category into validity composites improved signal detection over univariate cutoffs.

Comprehensive Neuropsychological Assessment of Concussion (Post-Injury Testing)

The goal of concussion testing for athletes is to determine when and if a player has returned to their pre-injury cognitive baseline and can return to sport participation. Test results are used in conjunction with other information (e.g., symptom endorsement, physical performance measures such as balance) to minimize the chance of an athlete sustaining another concussion before recovering from the first, with a potentially catastrophic outcome such as second impact syndrome (Saunders & Harbaugh, 1984). Notably, the results of a standalone computerized cognitive assessment such as the ImPACT are not designed to *diagnose* a concussion and should not be mistaken as equivalent to a comprehensive neuropsychological evaluation.

According to The International Concussion in Sport Group (CISG; McCrory et al., 2017), neuropsychological assessment is the cornerstone of concussion management and should be a part of any return-to-play protocol following a sport-related concussion. Neuropsychological testing increases diagnostic accuracy in the evaluation of sport-related concussion recovery (van Kampen et al., 2006) and can also identify ongoing functional changes in the absence of symptoms (Lovell et al., 2004; Fazio et al., 2007). Due to the inherent incentive to report being symptom-free (McCrea et al., 2004), this is an asset since athlete symptom self-report has long been considered to be unreliable (Field et al., 2003). As such, neuropsychologists often contribute to decision making regarding whether an athlete is ready to return-to-play following a concussion. Temple (2019) describes that these neuropsychological evaluations typically last three or more hours and can include the assessment of the domains of global intellectual functioning, academic abilities, attention, executive functioning, language, sensorimotor processing and functioning, visuospatial skills, memory, and emotional functioning. If neuropsychological deficits are identified, recommendations can be made regarding potential interventions, as well as necessary accommodations to the classroom or workplace.

Iverson and Schatz (2014) conclude that the value of neuropsychological assessment in the management of sport-related concussion has a strong empirical foundation. There are two general approaches: to complete neuropsychological testing only after an athlete's symptoms have resolved, or to complete testing shortly following the concussion. The first approach is often used as part of a stepwise process for clearing an athlete to return to sports and is generally more practical given the time and cost of a brief neuropsychological assessment. Additionally, Iverson and Schatz (2014) point out that scheduling repeated evaluations can be difficult and practice effects associated with repeated testing can make results more difficult to interpret. Some evidence has emerged, however, that early testing could have value for predicting recovery time (Iverson, 2007; Iverson et al. 2007). On the other hand, the second approach, has its benefits as well, such that testing within a short time period following a concussion might be useful to assist with early management recommendations. Iverson and Schatz (2014) provide the example of an athlete who may be unsafe to drive and would benefit from a greater duration of rest and activity limitations. Whereas on the other hand, an athlete who appears cognitively intact with only mild symptoms may be encouraged to engage in more activities as tolerated. Therefore, brief evaluations, while symptomatic, can be used to monitor recovery and to make recommendations regarding activity restrictions and academic accommodations.

Post-injury, some cases of concussion can involve symptoms that appear to be maintained by factors other than those directly related to the neurological injury. For example, Meares et al. (2011) found that premorbid anxiety and depressive disorders, as well as acute post-traumatic stress, were early markers for development of a post-concussion syndrome. Iverson and Lange (2011) concluded that post-concussion syndrome can in fact be worsened by psychological distress, social psychological factors (e.g., the nocebo effect, iatrogenesis, and misattributions), personality characteristics, and co-occurring conditions (e.g., chronic pain and insomnia). Several studies (Ferguson et al., 1999; Mittenberg et al., 1992) have suggested that prolonged post-concussion syndrome results from an underestimation of pre-injury incidence of symptoms and a re-attribution of these symptoms to the brain injury. One framework described by Mittenberg and Strauman (2000) describes that (1) typical symptom expectancies are activated when mTBI occurs and symptom expectancies can bias selective attention to internal states; (2) naturally occurring premorbid symptoms are attributed to mTBI, and the selective attention to the inherent stress of the trauma subjectively magnifies these symptoms; and (3) symptom expectancies are confirmed, and anxiety about the significance of symptoms maintains selective attention.

Estimating Premorbid Functioning in Athlete Populations

Of particular importance in neuropsychological assessment of sport-related concussion is the accurate estimation of an athlete's premorbid level of cognitive functioning, in the absence of or in conjunction with baseline test scores. Accurate estimation of premorbid functioning in athletics is crucial because incorrect estimation could result in an athlete prematurely returning to play before their concussion has resolved, a decision that risks the occurrence of the rare, but devastating outcomes such as second-impact syndrome (Saunders & Harbaugh, 1984), or being withheld from play for longer than is necessary.

Rabinowitz and Arnett (2012) evaluated the use of the WTAR as an estimate of premorbid ability in a sample of healthy college athletes and found that for those with estimated FSIQs greater than 107, the discrepancy between actual neuropsychological test scores and the WTAR FSIQ estimate was greatest. They then looked at athletes who went on to sustain a concussion and found that athletes with higher IQ estimates had a WTAR estimate obtained post-concussion that suggested greater post-concussion decline than that indicated by comparison with actual baseline neuropsychological performance. Stamm et al. (2015) suggest that word-reading ability may be an inappropriate indicator of premorbid intellect in American football players and instead should be used as an outcome variable since early exposure to head trauma could limit acquisition of word reading skills.

Asken et al. (2020) collected data regarding subclinical head impacts throughout early life, suggesting that this may stunt acquisition of word reading skills. The effect of concussion and early repetitive head injury exposure history on cognition across the lifespan cannot be characterized accurately without appropriate consideration of premorbid functioning and sociodemographic variables, and Houck et al. (2018) state that SES, race, and medical history beyond exposure to brain injury or subclinical brain trauma are important factors when interpreting variability in cognitive scores among collegiate athletes. Asken et al. (2020) found that the WTAR appears to be unrelated to history of self-reported concussions and/or repetitive subclinical head trauma exposure in collegiate athletes, and that sociodemographic and academic variables should be incorporated in test score interpretations for diverse populations like athletes.

One study recently reported that African American athletes and athletes from low SES backgrounds scored worse on baseline neurocognitive testing, and baseline memory and speed scores are significantly influenced by age and race, as well as a history of a neurodevelopmental disorder (Houck et al., 2018). These authors also stress that sport-specific differences in the proportional representation of various demographic variables (e.g., SES and race) may also be an important consideration within the broader biopsychosocial attributional model. As such, the estimation of premorbid functioning in diverse groups is a complicated matter with many potential considerations.

Estimating Premorbid Functioning in Diverse Groups

There are many potential problems with current methodologies for estimating premorbid intelligence on Western tests for those with diverse cultural backgrounds. As such, a multifaceted approach is suggested, whereby Fujii (2017) explains that clinicians should apply Lezak et al. (2012)'s best performance approach while also taking into the account the client's life experiences. This involves three steps: (a) obtain an initial estimate, (b) corroborate this estimate with additional data, and (c) adjust the initial score, if appropriate. Within this framework, Fujii (2017) proposes two methods of determining a valid estimate of premorbid functioning. The first method includes estimating premorbid IQ scores of the individual and then adjusting based on corroborative information, whereas the second method involves obtaining an IQ estimate of the person's country of origin and then adjusting accordingly. For culturally diverse clients, however, standard Western strategies for estimating premorbid IQs can be problematic for several reasons. This is because their culture, acculturation to Western norms, life experiences, and conception of intelligence can significantly differ from those of individuals from the dominant culture, so formulas for predicting premorbid functioning may not necessarily be valid.

For diverse groups, the WTAR (now TOPF) is recommended because it is superior to the WRAT-4 for more educated and higher functioning individuals (Mullen & Fouty, 2014). The Barona Index is also another demographic measure with a modification that East Asians (Chinese, Japanese, and Koreans) and Singaporeans be included in the White demographic category because of the literature supporting comparable mean average level IQs (Lynn & Meisenberg, 2010). For specific countries, culture-specific word recognition reading tests are recommended for estimating premorbid IQ. Tests have been developed in the following countries with varying levels of predictive validity: Taiwan (Chen et al., 2009), Japan (Matsuoka et al., 2006), France (Mackinnon & Mulligan, 2005), New Zealand (Barker-Collo et al., 2011), Portugal (Alves et al., 2012), Sweden (Tallberg et al., 2006), Germany (Lehrl et al., 1995), Argentina (Burin et al., 2000), and Spain (Del Ser et al., 1997).

There are several caveats for using word recognition reading tests with culturally diverse clients. First, clinicians should be cautious when using tests validated on American samples with clients in other English-speaking countries because norms may not be equivalent. For example, Mathias et al. (2007) found that although generally comparable, the NART and WTAR produced some grossly inaccurate scores in an

Australian sample that were 30-36 points off actual IQ scores. Second, allowances should be made for foreign accents (Lezak et al., 2012).

Once an individual premorbid IQ estimate is obtained, the next step is to corroborate estimates with other data. Sources include behavioral observations and hobbies, academic data, and occupational data. Regarding behavioural observations, Fujii (2017) recommends that clinicians should be noting their clients' vocabulary and reading skills, and how articulate they are. Regarding academic data, education level is one of the best predictors of IQ (Crawford & Allan, 1997), as academic data has a moderate to strong correlation with IQ. However, weaknesses in using GPA to predict IQ include the strong impact of motivation and conscientiousness on GPA (Cheng & Ickes, 2009), as well as differences in curriculum rigor between academic institutions and adverse life experiences that can impact one's schooling. A more objective source for estimating premorbid IQ could be scores on standardized achievement tests, as generally high correlations between academic achievement tests and IQ have been reported (Frey & Detterman, 2004) which led several researchers to develop formulas for estimating IQ based on achievement test scores.

Occupation is another demographic variable that correlates highly with IQ (Crawford & Allan, 1997). Childhood IQ has been found to be predictive of occupation success in adulthood (Huang, 2001), and IQs of retired professionals and managers were 16 points higher than retired operatives/laborers (Ryan, Paolo, & Dunn, 1995). See Huang (2001) for a table of IQs based on adults aged 25-38 from 1983-1990 surveys. Special awards and achievements should also be considered (e.g., awards, rankings, positions, recognition within an occupation or academic setting, fellow status in an organization, rankings for intellectual activities such as chess).

Another strategy that Fujii (2017) describes is estimating IQ scores for the country of origin. This involves obtaining an estimate of IQ scores on Western tests from the client's country of origin and then adjusting scores based on the client's demographic information interpreted within the context of norms for that country. This strategy is most appropriate for immigrants that were educated primarily in their country of origin and for whom English is not their primary language. There are several methods for obtaining an IQ estimate of performance on Western tests: (a) data from the literature, (b) estimates based on regional IQs, (c) calculations from standardized academic tests, and (d) calculations from gross domestic product per capita (GDP). This can be difficult, however, as most countries do not have Wechsler-based IQ data. Instead, many use a version of Raven's Progressive Matrices (Raven et al., 2003), and scores for other countries may be based on something as simple as the Draw-A-Person test. One comprehensive source for procuring IQ scores on Western tests from different countries is Lynn and Vanhanen (2002, 2006; Lynn & Meisenberg, 2010), who obtained IQ data from 81 countries and estimated IQs for another 104 nations based on scores for neighboring or ethnically similar countries. However, this dataset is highly controversial and has been heavily criticized for its methodology.

Fujii (2017) recommends that next clinicians should seek additional data to help corroborate and adjust IQ score estimates. Once a national IQ estimate is determined, the next step is to adjust the score for the client based on personal demographic data interpreted within the context of country of origin. Clinicians should determine quantity and level of education, occupation, reason for and timing of immigration, educational and occupational achievement of progeny, and English skills. Another pertinent indicator for estimating intelligence on Western tests is quality of education (Manly et al., 2002).

Occupation and socioeconomic status, both in the country of origin and in the country of residence, would be another indicator for determining premorbid IQ (Fujii, 2017). Additionally, the wave in which a person emigrates can also be an indicator of premorbid functioning (Fujii, 2017). The general pattern is that the first wave of immigrants includes the educated, professional, or affluent, and then others arrive in subsequent waves, many for family reunification. Priority workers are persons with outstanding abilities in the science, arts, education, business, or athletics; outstanding professors, researchers, or managers; or executives of multinational companies. Another consideration is the individual's children's academic achievement or occupation. This assumption is based on the moderately high correlation between the IQ of parents and children (.48; Jencks et al., 1972).

Neurodevelopmental, Psychiatric, and Medical Considerations

Consideration of a history of learning disability is important because premorbid functioning is often assessed with tests of word reading, which can be affected in adults with a history of special education service for reading difficulty (Semrud-Clikeman & Fine, 2008). There is also some evidence that histories of learning disability and mild TBI are independently related to lower baseline cognitive performance in college athletes (Collins et al., 1999). As well, students who report comorbid histories or histories of academic difficulties alone produce lower ImPACT composite scores, and those with comorbid histories or histories of ADHD alone produce invalid protocol warnings more frequently than student athletes without such histories (Manderino & Gunstad, 2018). Therefore, student athletes with a history of ADHD or academic difficulties may more frequently fall below validity score thresholds, suggesting caution in interpreting test performance on the ImPACT and other cognitive tests administered at baseline. Notably, it is possible that these genuine cognitive difficulties, either in cognitive domains assessed by validity subtests (e.g., attention) or inability to follow test instructions (e.g., learning disorder diagnosis or a history of poor academic performance), are suppressing ImPACT test scores, for some athletes, and thus increasing the rate of invalid protocol warnings. However, distinguishing genuine cognitive impairment from invalid performance is complicated, and comparison to normative data alone may not accurately depict postinjury impairments for student athletes with these conditions.

Overall, there is controversy in the literature about the degree to which psychological distress can affect cognitive test performance. Cognitive impairment has been widely reported in people who have major depression (Austin et al., 2001; Goodwin, 1997). Significant correlations between depression severity and cognitive performance have been found in the domains of episodic memory, executive function, and processing speed, but depression is less likely to impact domains such as semantic memory or visuo-spatial memory (McDermott & Ebmeier, 2009). Khan-Bourne and Brown (2003) suggest that there is a significant impact of depression on outcomes in persons with TBI or stroke. On the other hand, Sherman et al. (2000) found evidence for only a small effect of depression on neuropsychological test performance in individuals with TBI. Neither of these studies, however, considered the potential impact of degree of depression at the time of the evaluation in the context of protective factors like cognitive reserve or risk factors like financial compensation-seeking.

Individuals who are in an anxious state can also demonstrate difficulties on cognitive testing (Eysenck, 1992). The effects of anxiety on cognitive processing often center on the central executive component of Baddeley's (1986, 2001) working memory system. Cognitive deficits have been widely recognized to be an important component of anxiety. Specifically, anxiety is thought to restrict the capacity of working memory by competing with task-relevant processes. A meta-analysis by Moran (2016) found that self-reported measures of anxiety are reliably related to poorer performance on measures of working memory capacity, including complex span, simple span, and dynamic span tasks. They also noted that a review of the literature found that anxiety is related to poorer performance across a wide variety of tasks including fluid cognition more generally, but memory, attentional control, processing speed, reasoning, and vocabulary more specifically.

In summary, there are many factors that contribute to the estimation of premorbid cognitive functioning, including the test battery and score implications, norms used, demographic and personal variables, age of the examinee, and comorbid disorders and diagnoses. Therefore, it is important to be aware of these factors and consider their impact for each person seen for a neuropsychological assessment, as outlined by the AACN's recommended "precision medicine" approach. At the current time, neuropsychological assessment of sport-related concussion is lacking in this knowledge base and in approaches that emphasize the importance of the consideration of these factors in estimating premorbid functioning.

The Current Study

This research project aims to improve upon the estimation of premorbid cognitive functioning in varsity athletes who have sustained a sport-related concussion. Accurate estimation of premorbid functioning in athletics is crucial because incorrect estimation could result in an athlete prematurely returning to play before their concussion has resolved, being withheld from play for longer than is necessary, or sustaining devastating outcomes such as second-impact syndrome (Saunders & Harbaugh, 1984). As such, this research benefits athletes in that it aims to improve concussion safety.

Improving methods of estimating premorbid cognitive functioning in this population is also important because baseline testing is a resource-intensive endeavor with many shortcomings that could eventually be replaced or shortened thanks to more precise actuarial approaches. A goal of this study is to provide rich and novel information regarding best estimates of premorbid functioning in sport neuropsychology, taking a deeper look at and considering important demographic variables when making return-to-play decisions, with the aim of taking a step towards eliminating baseline testing. This study also takes a more specific test score and domain-based approach to estimating premorbid functioning (i.e., estimating specific aspects of cognition), rather than the approach of much of the past research which has attempted to estimate IQ score.

The preceding literature review demonstrates the importance of the accurate determination of athletes' cognitive return to baseline, highlighting the impact that improper techniques can have regarding concussion. In keeping with the AACN's consensus statement, this study ultimately aims to define expected performance more accurately through the incorporation of a broader range of variables in prediction models, thus taking more of a “precision medicine” approach.

Little improvement has been made to methods of estimating premorbid cognitive functioning in clinical neuropsychology in recent years, and much of the past research has been with older adults or people who have sustained more severe acquired brain injuries. In athletics, using a direct approach to the measurement of premorbid functioning through baseline testing is resource intensive and difficult, due to issues of performance validity. It is often still used, however, because this population is at such a high risk of sustaining a concussion. Many baseline assessments also occur in group settings, which opens the door for more distraction than in a typical testing environment. Abeare et al. (2018) found that the base rate of failure of performance validity indicators on the ImPACT during baseline assessment was surprisingly high overall. In addition, the concept of “sandbagging” a baseline assessment has become quite popular, whereby athletes intentionally perform poorly to artificially lower their baseline score to return-to-play post-concussion faster. Therefore, indirect measurements of premorbid functioning may be more telling, and work must be done to strengthen the utility of these methods in sport neuropsychology. Thus, the current research project will evaluate the efficacy of several methods of estimating premorbid functioning in an athlete population and attempt to develop a regression equation in which demographic variables and test scores are used to predict premorbid functioning via specific test scores and domains of cognition rather than through more general IQ estimates.

Study I: Identifying Variables that Best Estimate Premorbid Functioning

The first study is a general exploration of the measures and variables that best relate to measures deemed to be useful in the estimation of premorbid functioning, specifically in a varsity athletics setting. Thus, this study aims to evaluate the relationships between well-established measures of premorbid functioning, as well as tests hypothesized to add to our understanding of estimating premorbid functioning (e.g., verbal fluency), and a wide range of demographic and personal information collected. These demographics go beyond age, sex, and race, including detailed educational data, geographics, parental data, and occupational data (see Appendix C).

These sociodemographic variables are expected to be correlated with performance on well-established measures of indirect assessment of premorbid functioning, including the TOPF. It was hypothesized that the strongest relationships will be found between tests of crystallized intelligence and educational demographic variables, including the TOPF and likely phonemic fluency, as well as years of education, quality of education, and parental education. Therefore, the aim is to replicate and expand upon our understanding of these relationships.

Study II: Using Verbal Fluency Word Quality as a Predictor Variable of Level of Premorbid Functioning

Preliminary research by Abeare and Seguin (2014) found that people with a higher NAART estimated IQ are more likely to generate words on a phonemic fluency task (FAS) with a higher maximum Age of Acquisition (AoA). Therefore, Study 2 aims to explore and extend upon Abeare and Seguin (2014)'s findings, to determine whether the quality of athlete responses on the Delis-Kaplan Executive Function System

(D-KEFS) Verbal Fluency Test - Phonemic Fluency is related to estimates of premorbid functioning provided by the TOPF. It is hypothesized that much like in Abeare and Seguin's (2014) study, word qualities related to the AoA of the words produced will be a strong predictor of premorbid functioning and will account for unique variance in the prediction of premorbid functioning. If this metric correlates well with baseline cognitive performance, it will be added as a predictor variable in the creation of the regression equation outlined in Study 3 to assess the incremental validity of this metric.

Study III: A Regression-Based Approach to Estimating Premorbid Functioning

The third study seeks to examine whether it is possible to use a unique combination of the demographic variables and test scores identified to be most indicative of one's premorbid level of functioning in Study 1 and 2 to be able to develop a regression-based approach to the estimation of premorbid functioning in sport neuropsychology. The aim is to generate prediction algorithms using standard multiple regression in which subtest raw scores and demographic variables are predictors of ImPACT composite scores and other neuropsychological tests that are commonly used in sport concussion assessment (i.e., Verbal Fluency, Trails A and B), unlike past research which has largely aimed to estimate IQ scores. Neuropsychology as a field should continuously improve upon its procedures based on new evidence, and must explicitly identify, define, and measure the disparities associated with demographic variables to quantify their impact on test performance (AACN, 2021). The development of testing methods and practices that reduce bias and inequity in clinical assessment and decision-making are of particular concern in sport-related concussion assessment and management, as concussion is a complex issue, and athletes are a diverse group of individuals.

It is hypothesized that the use of regression-based methods using multiple relevant predictor variables could improve the accuracy of premorbid functioning estimates beyond current approaches, while also emphasizing cultural competency in assessing diverse populations by incorporating a broader range of variables to create more precise and personalized estimates specific to this unique population. This approach also aims to help reduce clinician biases and how they intersect with client identity. If successful, this approach could be used as a supplement to shorter baseline assessments or could act as the first step in eventually replacing baseline testing entirely.

CHAPTER 2

General Method

Participants

158 athletes (M_{age} = 20.30, SD_{age} = 1.95, range = 17-25; 70% male; 64% White; M_{edu} = 13.20, SD_{edu} = 1.44) were recruited as a part of the University of Windsor's varsity athletics baseline testing effort from the following sports: men and women's basketball (n=16; n=13), men and women's hockey (n=2; n=13), men and women's soccer (n=1; n=8), men and women's volleyball (n=7; n=13), and men's football (n=85). All athletes on these sport teams (which are considered to be high impact and thus have a higher concussion probability) apart from those who had recently (within the past six months) sustained a concussion and completed a post-injury assessment were eligible to participate. Additionally, those who failed performance validity tests were excluded, as Stosic et al. (2024) conclude that failing to exclude data from carelessly responding

participants on questionnaires or behavioural tasks frequently results in false-positive or spuriously inflated findings. More specifically, athletes who obtained a score of less than 45 on the Word Choice Test (Tyson et al., 2023; n = 3), who failed the ImPACT's Default EVI (Lovell, 2016; n = 3), who failed the ImPACT RedFlags EVI (Lovell, 2011; n = 4), or who failed the ImPACT 5 A criteria (Erdodi et al., 2020; n = 8) were removed (total n = 13, some athletes failed multiple criterion).

Measures

Each athlete completed a fixed battery of tests, beginning with consent and a pre-testing questionnaire; See Appendix A for the complete battery list and order, Appendix B for the consent form, and Appendix C for a summary of the contents of the pre-testing questionnaire. For the purposes of this study, only the following neurocognitive performance-based tasks are included:

Immediate Post-Concussion Assessment and Cognitive Testing – Version 4 (ImPACT-v4)

The ImPACT (Lovell, 2022) is a computerized neurocognitive test that measures aspects of memory, attention, visual spatial processing, impulse control, and processing speed in individuals from 12 to 80 years of age. The test begins with the collection of demographic information, followed by a self-report concussion symptom scale. The neurocognitive test modules are then administered in the following order: Word Memory, Design Memory, X's and O's, Symbol Match, Color Match, Three Letters, Word Memory Delayed Recall, and Design Memory Delayed Recall (see Appendix D for a description of each subtest). Test administration is typically completed within 20 minutes, and all scoring is automatically completed by the software. The standardization sample consisted of 72,369 individuals who completed baseline ImPACT testing. In addition to specific scores that are provided for each module, the following composite scores are also reported: Verbal Memory, Visual Memory, Visual Motor Speed, Reaction Time, and Impulse Control. A Total Symptom Composite Score is also provided.

Test of Premorbid Functioning (TOPF)

The Test of Premorbid Functioning (TOPF; Wechsler, 2011) is a revised and updated version of the Wechsler Test of Adult Reading (WTAR; Holdnack, 2001) and it is used to obtain an estimate of individual's premorbid cognitive and memory functioning. It includes a list of 70 words that have atypical grapheme to phoneme translations and requires the athlete to read each word out loud with proper pronunciation. The TOPF was co-normed with the Wechsler Adult Intelligence Scale – Fourth Edition (WAIS-IV; Wechsler, 2008).

Delis-Kaplan Executive Function System – Verbal Fluency Subtest (D-KEFS VF)

The D-KEFS Verbal Fluency Test (Kaplan & Kramer, 2001) is comprised of three testing conditions: Letter Fluency, Category Fluency, and Category Switching. For the Letter Fluency condition, the athlete is asked to generate words that begin with a particular letter as quickly as possible. For the Category Fluency condition, the athlete is asked to generate words that belong to a designated semantic category as quickly as possible. The last condition, Category Switching, requires the examinee to generate words, alternating between two different semantic categories as quickly as possible. Two scores are obtained for this condition, the number of words correctly generated, and the switching accuracy score (i.e., did they alternate properly between categories). For each trial of each condition, the examinee is allowed 60 seconds. This test measures the examinee's ability to generate words fluently in an effortful, phonemic format (Letter Fluency), from overlearned concepts (Category Fluency), and while simultaneously shifting between overlearned concepts (Category Switching). All athletes were administered the Standard Form. The D-KEFS was standardized on a nationally representative, stratified sample of 1750 children, adolescents, and adults, ages 8-89 years. The age groups relevant to this study included 175 people ages 16-19 years, and 175 people ages 20-29, and were roughly equal in terms of sex composition.

Trail Making Test (TMT) A and B

The Trail Making Test (TMT; Reitan, 1955) is a neuropsychological test that is commonly used to assess executive functioning, attention, processing speed, and visual motor skills. Heaton norms were used for the TMT, once again correcting for age, education, and Ethnicity (Heaton et al., 2004). TMT A presents individuals with a page of randomly dispersed numbers and asks the athlete to connect them with a line, in order, as quickly as possible, and on TMT B numbers and letters are dispersed together, and individuals are asked to alternate from numbers to letters, in order.

Procedure

Baseline testing efforts were completed between April 1st and December 1st, 2022. Athletes who participate in high contact sports (i.e., football, soccer, basketball, volleyball, hockey) and who had not yet completed baseline testing were recommended to participate by their team coaches and athletic trainers. Before completing the battery of cognitive tests (Appendix A), athletes completed a questionnaire online, which began with consent (Appendix B). After completing the consent process, namely that they agreed to undergo a sports concussion evaluation and agreed that their de-identified data could be used for teaching and research purposes, they completed a broad range of questions that took approximately 20 minutes to complete. The questionnaire collected detailed demographic information, questions from the Depression Anxiety Stress Scales (DASS), the Difficulties in Emotion Regulation Scales (DERS), questions about their developmental and medical history, and the Athlete Sleep Screening Questionnaire (ASSQ), but these measures were not included in this dissertation. An outline of the questions asked within the comprehensive questionnaire completed can be found in Appendix C. Athletes completed all tests within the in-person battery in the order denoted in Appendix A.

CHAPTER 3

Study I: Identifying Variables that Best Estimate Premorbid Functioning

Level of premorbid cognitive functioning is estimated by comparing current performance with an estimate of original ability level and is defined as an indirect measurement procedure (Lezak et al., 2013). There are several methods used for doing so, including historical and observational data, mental ability scores, word reading tests, demographic variable formulas, and test score and demographic hybrid formulas (Lezak et al. 2012). Age, sex, race, education, and occupation have commonly been shown to have a strong relationship with IQ (Matarazzo, 1972). Proxy measures of cognitive reserve, such as socioeconomic status, are often considered as well. Currently, no known studies have included a broader and more specific range of factors like achievement, parental education, and family income (see Appendix C for a comprehensive list) when seeking to improve the estimation of premorbid functioning.

Study Aims and Hypothesis

The current study's aim was to explore which demographic and personal variables best correlate with athlete performance on cognitive testing, using a broader and more detailed range of variables in this investigation (see Appendix C). A range of cognitive tests were included, including those commonly used in sport-related concussion assessment, such as the ImPACT. It was hypothesized that the strongest relationships would be found between tests measuring aspects of crystallized intelligence (i.e., the TOPF) and educational demographic variables such as years of education, GPA, parental education, and academic scholarship attainment, as well as variables related to socioeconomic status. It was also hypothesized that performance on verbal fluency composite scores would be related to performance on the TOPF as well, in keeping with Abeare and Seguin's (2014) findings related to the NAART.

Method

Data Analysis

Preliminary analyses. All analyses were conducted using IBM SPSS Statistics v.29.0.0.0. Both raw and normed test scores (i.e., composite, standard scores, scaled scores, and T scores) were used: raw scores were used because they do not possess the influence of a normative sample, and normed test scores were used to incorporate the influence of each test's demographic corrections. Prior to conducting the primary analyses, the assumptions of correlation and ANOVA analyses were tested. Correlations were calculated to examine the relationships between continuous and dichotomous variables in order to help determine if changes in one variable are associated with changes in another variable. ANOVA analyses were conducted when evaluating the relationship (i.e., mean differences) between nominal/categorical variables and continuous variables, as there are no correlation coefficients that are appropriate for this use. Assumptions of ANOVA include normality of variables (e.g., histograms, Q-Q plot, Kolmogorov-Smirnov test), equal variances of the populations that the samples come from (e.g., boxplots, Bartlett's test), and independence of observations. Notably, there were significant differences in the number of athletes within several of the groups that were to be evaluated using ANOVA, and thus the variance was not always equal. See Table 1, 2, 3, 4, and 5 for a breakdown of these discrepancies for variables that were categorical in nature. For the variables outlined in Tables 1-5, Levene's test was used to determine homogeneity of variance across each dependent variable, and filtering was used to include the groups with the largest sample sizes if Levene's test was violated. Additionally, some variables, such as postal code and parent occupation, when categorized, resulted in most variables containing only one response. Running an ANOVA with a group containing only one observation is not statistically valid, since ANOVA relies upon comparing means across groups. Because of this, the influence of these variables was considered to be accounted for and more accurately captured by family income and impression of SES (for postal code) and number of years of parent education and family income (for parent occupation).

Table 1.

Number of Athletes by Racial Category

	White	Black	Asian	Hispanic	Indigenous	Other
Race	101	35	6	6	3	7

Table 2.

Number of Athletes by Sport

	Men's Basketball	Men's Football	Men's Hockey	Men's Soccer	Men's Volleyball	Women's Basketball	Women's Hockey	Women's Soccer	Women's Volleyball
Sport	16	85	2	1	7	13	13	8	13

Table 3.

Number of Athletes by Major Area of Study

	Kinesiology	Business Administration	Criminology	Sport Management and Leadership	Psychology	Nursing	Other
Major	31	25	12	11	9	5	65

Note. Most major areas of study had less than 5 people in them and for the purposes of this table, are included in the "Other" category.

Table 4.

Number of Athletes by First Language

	English	French	English and Arabic	English and French	English and Other	Other
First Language	133	10	1	3	4	3

Table 5.

Number of Athletes by Birth Country

	Canada	USA	Nigeria	Ivory Coast	Australia	Brazil	Cameroon	Dubai	Ethiopia	Turkey	Venezuela	Prefer not to Answer
Birth Country	127	11	3	2	1	1	1	1	1	1	1	7

Assumptions of correlation analyses include level of measurement (i.e., variables should be measured at the interval or ratio level for Pearson correlations; if variables are ordinal and if the data is non-normal, Spearman correlations are used; if one variable is dichotomous, Point-Biserial correlations are used), linear relationship (e.g., scatterplots), normality of variables (e.g., histograms, Q-Q plot), related pairs (e.g., each observation has a measurement for each variable), and absence of outliers (e.g., Cook's distance, Mahalanobis distance). Because it is common for normality to be violated in these types of data, the use of Spearman correlations instead of Pearson correlations is justified (Bishara & Hittner, 2015). In keeping with Dancey and Reidy (2004), correlations between 0 and ±0.3 were considered to be weak, correlations between ±0.4 and ±0.6 were considered to be moderate, and correlations between ±0.7 and ±0.9 were considered to be strong. Perfect correlations hold a value of ±1.0.

Results

The descriptive statistics for each test score are found in Table 6. A more comprehensive and descriptive list of athletes' major area of study, race, first language, birth country, and mean test scores by sport type is found in Appendix E.

Table 6.

Descriptive Statistics for Cognitive Test Scores

Test Score	n	*Raw Score* *M* (SD)	Standardized Score *M* (SD)	Range Raw; Range SS
ImPACT Verbal Memory Composite Score	158	88.38 (8.71)		62-100
ImPACT Visual Memory Composite Score	158	76.26 (11.19)		45-99
ImPACT Visual Motor Speed Composite Score	158	40.88 (5.58)		22.45-53.25
ImPACT Reaction Time Composite Score	158	0.60 (0.09)		0.46-1.12
ImPACT Impulse Control Composite Score	158	5.00 (3.57)		0-18
TOPF Scores	158	41.47 (9.92)	SS = 104.18 (11.05)	15-63; 76-127
Trail Making Test A Scores	158	23.33 (7.28)	T = 51.46 (11.35)	11-48; 29-87
Trail Making Test B Scores	157	59.81 (28.39)	T = 51.31 (11.17)	11-219; 18-80

Verbal Fluency Letter Fluency Scores	157	38.30 (9.68)	ss = 10.78 (3.01)	4-69; 1-19
Verbal Fluency Category Fluency Scores	157	45.46 (9.44)	ss = 13.07 (3.57)	18-79; 2-19
Verbal Fluency Category Switching Number Correct Scores	157	14.24 (2.65)	ss = 11.16 (3.17)	6-20; 2-18
Verbal Fluency Category Switching Accuracy Scores	157	13.32 (2.89)	ss = 11.77 (2.87)	6-21; 5-19

Note. One participant did not complete their assessment in-full. SS = Standard Score; T = T-score; ss = scaled score.

The Test of Premorbid Functioning (TOPF)

The TOPF raw and Standard Score (SS) values were correlated with the demographic and personal variables outlined in Appendix C and other objective measures (i.e., TMT A and B, ImPACT composite scores, D-KEFS VF) using Spearman correlations (when correlating with continuous variables) and point-biserial correlations (when correlating with dichotomous variables). TOPF raw score and Standard Score were significantly and weakly correlated with several demographic and personal variables (see Table 7).

Table 7.

TOPF Raw and Standard Score Correlations with Demographic and Personal Variables

	Years of Education	Age	GPA	Grade 7/8 Math	Grade 7/8 Language	Grade 7/8 Science	Grade 7/8 Social Studies	Grade 7/8 Art	Parent 1/Mother Education	Parent 2/Father Education	Birth Order	SES Ladder	Family Income
TOPF Raw Score	.121	.141	.085	**.163***	.093	.053	.142	.022	.080	**.174***	.057	.020	.087
TOPF SS	.028	.013	.111	**.212****	.139	.093	**.172***	.051	.105	**.181***	.059	-.024	.110

Note. *p <.05, **p < .01. A two-tailed bivariate Spearman correlation was conducted to account for non-normal distributions. Significant correlations with TOPF raw score included Grade 7/8 math grades and father/parent 2's education, and significant correlations with TOPF standard score included grade 7/8 math, grade 7/8 social studies, and father/parent 2's education.

Regarding objective cognitive testing, TOPF raw and Standard Scores were weakly and positively correlated with several D-KEFS Verbal Fluency scores (see Table 8).

Table 8.

TOPF Raw and Standard Score Correlations with Objective Cognitive Test Scores

	ImPACT Verbal Memory	ImPACT Visual Memory	ImPACT VMS	ImPACT RT	ImPACT IC	TMT A Raw	TMT A T-Score	TMT B Raw	TMT B T-Score	VF-LF Raw	VF-LF ss	VF-CF Raw	VF-CF ss	VF CS # Raw	VF CS # ss	VF CS Acc Raw	VF CS Acc ss
TOPF Raw Score	-.018	-.003	.082	.044	.064	-.026	.019	-.089	.08	**.162***	**.159***	.147	.152	**.162***	**.164***	.128	.105
TOPF SS	-.043	-.010	.078	.047	.082	-.034	.036	-.108	.096	.140	**.161***	.142	.149	.146	**.161***	.111	.111

Note. *p <.05, **p < .01. A two-tailed bivariate Spearman correlation was conducted to account for non-normal distributions. Significant correlations to note include VF-LF raw score, VF-LF scaled score, and VF-CS number correct raw and scaled score.

Point-biserial correlations were conducted for dichotomous variables, with weak positive correlations between TOPF raw and Standard Score and dichotomous variables (see Table 9). Specifically, a higher TOPF raw or SS was related to the absence of a specific learning disorder diagnosis in written expression, and a higher TOPF raw score was correlated with part-time studies.

Table 9.

TOPF Raw and Standard Score Correlations with Dichotomous Variables

	Sex	Handedness	Full/Part Time Studies	Scholarship (Yes/No)	SLD Read	SLD Write	SLD Math	ASD	Speech Disorder	ADHD	Other
TOPF Raw Score	-.055	-.012	**.162***	-.033	-.153	**-.190***	-.045	.053	-.036	.063	.090
TOPF SS	-.076	-.022	.144	-.050	-.144	**-.163***	-.022	-.045	-.044	.079	.096

Note. *p <.05. A point-biserial correlation was conducted using bootstrapping to account for non-normal distributions. Significant correlations include TOPF scores with full vs. part-time studies, and SLD in written expression (a higher TOPF raw or SS was related to the absence of an SLD in written expression and part-time studies).

One-way ANOVA analyses were conducted when evaluating mean differences between nominal/categorical variables (independent variable/factor) and the continuous TOPF variables (dependent variable), and Levene's test was used to test for homogeneity of variances. These variables included race, sport, major area of study, previous degree, first language, and birth country. If violated, Levene's test is commented on for each variable.

There was heterogeneity of variance for race, so a filter was applied to include the racial categories with the most athletes, namely White (n=101) and Black (n=35) students. This time, there was homogeneity of variances for TOPF raw and SS for White and Black students. A one-way ANOVA was conducted for Racial category (including only White and Black athletes), and was not significantly related to TOPF raw score, $F(1,134) = .02$, $p = .88$, or TOPF SS, $F(1, 134) = .24$, $p = .63$.

A one-way ANOVA was conducted and sport was not significantly related to TOPF raw score, $F(8, 149) = 1.55$, $p = .15$, or TOPF SS, $F(8, 149) = 1.57$, $p = .14$. A filter was applied to remove those who played men's hockey (n=2) and men's soccer (n=1) due to small sample sizes, and sport was still not significantly related to TOPF raw score, $F(6, 148) = 1.62$, $p = .15$, and TOPF SS, $F(6, 148) = 1.54$, $p = .17$. Despite sample size across most groups being quite small and unequal, homogeneity of variance existed across the major area of study variable with regards to TOPF raw and SS ($p > .05$). Major area of study was not significantly related to TOPF raw score, $F(42, 115) = 1.15$, $p = .28$, nor was it related to TOPF SS, $F(42, 115) = 1.14$, $p = .29$. A filter was applied to investigate the relationship between the major areas of study that were most common among athletes, including Kinesiology (n=31), Business Administration (n=25), Criminology (n=12), Sport Management and Leadership (n=11), Psychology (n=9), and Nursing (n=5). Major area of study including these majors only was not significantly related to TOPF raw score, $F(4, 73) = .89$, $p = .47$, or TOPF SS, $F(4, 73) = .80$, $p = .52$.

First language did not significantly relate to TOPF raw score, $F(6,151) = .96$, $p = .48$, nor did it relate to TOPF SS, $F(6, 151) = .95$, $p = .46$. A filter was applied to include only the first languages with the most athletes, namely English (n=133) and French (n=10). Again, first language did not significantly relate to TOPF raw score, $F(1, 141) = .00$, $p = .99$, or TOPF SS, $F(1, 141) = .02$, $p = .89$.

Lastly, birth country did not significantly relate to TOPF raw score $F(14, 143) = .95$, $p = .50$, nor did it significantly relate to TOPF SS, $F(14, 143) = .97$, $p = .49$. A filter was applied to investigate the birth countries with the most athletes, namely Canada (n=127) and the USA (n=11). Birth country (including only Canada and the USA) did not significantly relate to TOPF raw score, $F(1, 136) = .15$, $p = .70$, or TOPF SS, $F(1, 136) = .31$, $p = .58$.

Immediate Post-Concussion Assessment and Cognitive Testing (ImPACT)

The relationship between the ImPACT verbal memory composite score and the demographic and personal variables outlined in Appendix C and other objective measures (i.e., TMT A and B, TOPF, D-KEFS VF) was investigated using Spearman correlations (when correlating with continuous variables) and point-biserial correlations (when correlating with dichotomous variables). TOPF correlations with the ImPACT composite scores can be found in Table 8.

Results of two-tailed bivariate Spearman correlations among ImPACT composite scores and demographic and personal variables are described as the following. The verbal memory composite score revealed only one significant positive but weak correlation, the visual memory composite score revealed two significant but weak correlations, the visual motor speed composite score revealed several significant weak positive correlations, and reaction time composite score revealed several significant weak correlations (see Table 10). Note that the impulse control composite score provides a measure of errors on testing, thus indicating the sum of errors committed during different phases of the test. As such, lower impulse control composite scores indicate better performance. Results of two-tailed bivariate Spearman correlations among impulse control composite scores and demographic and personal variables revealed no significant correlations (see Table 10).

Table 10.

ImPACT Composite Score Correlations with Demographic and Personal Variables

	Years of Education	Age	GPA	Grade 7/8 Math	Grade 7/8 Language	Grade 7/8 Science	Grade 7/8 Social Studies	Grade 7/8 Art	Parent 1/Mother Education	Parent 2/Father Education	Birth Order	SES Ladder	Family Income
Verbal Memory Composite	.108	.075	.147	.146	.076	.251**	.144	.111	.049	.105	-.003	.006	.025
Visual Memory Composite	-.078	-.056	.165	.210**	.102	.199**	.020	.074	-.014	.036	.036	.122	.003
Visual Motor Speed Composite	-.107	-.053	.208*	.270**	.174*	.226**	.201*	.039	.059	.017	-.012	.034	.040
Reaction Time Composite	.080	.070	-.297**	-.224**	-.166*	-.163*	.160	.006	-.137	-.012	.009	-.058	-.068
Impulse Control Composite	.084	-.009	-.066	.030	.050	.080	.032	.126	.004	-.008	.049	-.032	-.146

Note. *p <.05, **p < .01. Two-tailed bivariate Spearman correlation was conducted to account for non-normal distributions. Significant personal variables with correlations with ImPACT composite scores include: verbal memory and grade 7/8 science grades (higher verbal memory composite was related to higher self-reported grade 7/8 science grades), visual memory and grade 7/8 math and science grades (higher visual memory composite scores were related to higher self-reported grade 7/8 math and science grades), visual motor speed and GPA, grade 7/8 math, grade 7/8 language, grade 7/8 science, and grade 7/8 social studies (higher visual motor speed composite scores were related to higher GPA and grade 7/8 grades), and reaction time composite and GPA, grade 7/8 math, grade 7/8 language, and grade 7/8 science (lower reaction time is related to a higher GPA and grade 7/8 grades). No significant correlations were found for the impulse control composite.

Results of correlations between the verbal memory composite and other objective test scores revealed significant correlations with ImPACT visual memory and visual motor speed composite scores. Correlations between the visual memory composite and other objective test scores revealed significant correlations with other ImPACT composite scores as well as Trails B raw score. Correlations between the visual motor speed composite and other objective test scores revealed significant correlations with several other test scores, including the ImPACT reaction time composite, trail making test, and D-KEFS verbal fluency. Results of correlations between the reaction time composite and other objective test scores revealed significant correlations with the trail making test and D-KEFS verbal fluency tasks. Correlations between the impulse control composite and other objective cognitive test scores revealed one significant correlation with the visual memory composite. These results can all be found in Table 11.

Table 11.

ImPACT Composite Score Correlations with Objective Cognitive Test Scores

	Visual Memory Comp	VMS Composite	RT Composite	IC Composite	Trails A Raw	Trails A T-Score	Trails B Raw	Trails B T-Score	VF-LF Raw	VF-LF ss	VF-CF Raw	VF-CF ss	VF CS # Raw	VF CS # ss	VF Acc Raw	VF Acc ss
Verbal Memory Composite	**.469****	**.189***	-.040	**-.232****	.006	-.092	-.128	.066	.140	.116	.049	.070	.070	.031	.049	.018
Visual Memory Composite	**1**	**.175***	**-.161***	**-.181****	.082	-.127	**-.214****	.137	.135	.118	.016	.043	.038	.075	.006	.046
VMS Composite	**.175***	**1**	**-.547****	-.101	**-.363****	**.376****	**-.393****	**.382****	**.286****	**.301****	**.350****	**.363****	**.306****	**.264****	**.287****	**.254****
Reaction Time Composite	**-.161***	**-.547****	1	-.016	**.338****	**-.245****	**.345****	**-.311****	**-.221****	**-.226****	**-.344****	**-.340****	**-.261****	**-.244****	**-.286****	**-.284****
IC Composite	**-.181****	-.101	-.016	-.082	.088	.040	-.042	.123	.131	.118	.095	-079	.078	.082	.073	-.082

Note. *p <.05, **p < .01. A two-tailed bivariate Spearman correlation was conducted to account for non-normal distributions. Significant correlations between objective test scores and ImPACT composite scores can be seen in bolded text.

Dichotomous variables were also evaluated, and no significant correlations were present for verbal memory. One significant but weak correlation was identified between the visual memory composite score and students who had attained a scholarship and one significant but weak correlation was identified between the visual motor speed composite score and sex. One significant but weak correlation was identified between the reaction time composite score and sex (see Table 12).

Table 12.

ImPACT Composite Correlations with Dichotomous Variables

	Sex	Handedness	Full/Part Time Studies	Scholarship (Yes/No)	SLD Read	SLD Write	SLD Math	ASD	SLI	Speech Disorder	ADHD	Other
Verbal Memory Composite	-.090	.047	-.128	.046	-.086	-.045	-.001	.078	-.011	-.011	-.055	.031
Visual Memory Composite	-.068	-.002	.036	**.180***	-.082	-.106	-.137	.062	-.045	-.045	-.109	.034
VMS Composite	**-.213****	.027	.116	.091	-.030	-.079	-.097	-.104	-.127	-.127	-.028	-.022
Reaction Time Composite	**.175***	-.091	-.127	.137	.033	.020	.065	.008	-.006	-.006	.077	-.035
IC Composite	-.009	-.068	-.017	-.090	-.009	.029	.071	.023	.043	.043	.019	.050

Note. *p <.05, **p < .01 A point-biserial correlation was conducted using bootstrapping to account for non-normal distributions. Significant correlations between dichotomous variables and composite scores to note included sex (females tended to perform better than males on visual-motor speed and reaction time) and scholarship (those who performed well on visual memory were more likely to have a scholarship).

One-way ANOVA analyses were conducted when evaluating the mean difference between nominal/categorical variables (independent variable/factor) and the continuous verbal memory composite variable (dependent variable), as was done for the TOPF outlined above. Results did not change when analyses were conducted with a filter to only look at conditions with the most cases, except where indicated.

It was determined that race was not significantly related to verbal memory composite score, $F(5, 152) = .77$, $p = .57$. Levene's test indicated that there was heterogeneity within the verbal memory composite scores across major area of study ($p < .05$). A filter was applied to include only the major areas of study with the most athletes (i.e., Kinesiology, Business Administration, Criminology, Sport Management and Leadership, Psychology, and Nursing), and Levene's test was no longer significant ($p > .05$). As such, a one-way ANOVA was performed using these majors, and there was no significant difference in verbal memory composite across majors, $F(4, 53) = 1.16$, $p = .34$.

No significant difference existed between first language and verbal memory composite score, $F(6, 151) = 1.09$, $p = .37$. Birth country was also investigated, and one-way ANOVA revealed no significant difference between verbal memory composite score and birth country, $F(6, 151) = 1.09$, $p = .37$.

There were no racial differences for visual memory composite score, $F(5, 152) = .58$, $p = .71$. Sport was also not significantly related to visual memory composite score, $F(8,149) = .72$, $p = .67$. Major area of study was also not significantly related to the visual memory composite score, $F(42, 115) = .69$, $p = .92$. First language was also not significantly related to the visual memory composite score, $F(6, 151) = 1.42$, $p = .21$. Lastly, birth country was not significantly related to the visual memory composite score, $F(14, 143) = .59$, $p = .87$.

The visual motor speed composite score did not significantly differ dependent on race, $F(5, 152) = .26$, $p = .94$. The visual motor speed significantly differed dependent on sport, $F(8, 149) = 2.47$, $p = .015$. When the data was filtered to remove the two sports that had a low number of respondents, there continued to be a difference in the visual motor speed dependent on sport, $F(5, 142) = 3.74$, $p = .003$. The Tukey-Kramer post-hoc test was used due to unequal sample sizes across sports and indicated that the visual motor composite score was significantly greater for women's hockey (M = 44.14, SD = 6.31) than men's football (M = 38.86, SD = 5.88).

There was a statistically significant difference for major area of study, $F(42, 115) = 2.04$, $p = .001$. Given that there were so many major areas of study with only one respondent, this was re-conducted using the aforementioned major areas of study with the most respondents, and the results were no longer statistically significant, $F(5, 83) = 1.43$,

p = .22. First language was also investigated, and it was found that the relationship between the visual motor speed composite and first language was not statistically significant, $F(6, 151) = .96$, $p = .46$. Lastly, birth country's relation to the visual motor speed composite was investigated, and there was no statistically significant relationship, $F(14, 143) = .76$, $p = .71$.

There were no racial differences for the reaction time composite score, $F(5, 152) = .87$, $p = .50$. Sport was also not significantly related to reaction time composite score, $F(8,149) = 1.83$, $p = .08$, but it was statistically significant when the two sports with the smallest sample size were removed, $F(5, 142) = 2.72$, $p = .02$. The Tukey-Kramer post-hoc test was used due to unequal sample sizes across sports and indicated that the reaction time composite score was significantly lower for women's hockey (M = 0.54, SD = 0.05) than men's football (M = 0.62, SD = 0.09). Notably, a lower reaction time composite score indicates a faster reaction time, as this composite score evaluates the average response speed.

Major area of study was filtered to include only the majors with the highest number of respondents. Overall, major area of study (including only Business Administration, Kinesiology, Sport Management and Leadership, Criminology, and Psychology) was not significantly related to the reaction time composite score, $F(4, 79) = 1.76$, $p = .15$. First language was also not significantly related to the reaction time composite score, $F(6, 151) = 1.29$, $p = .27$. Lastly, birth country was not significantly related to the reaction time composite score, $F(14, 143) = .58$, $p = .87$.

There were no racial differences for the impulse control composite score, $F(5, 152) = .89$, $p = .49$. Sport was also not significantly related to the impulse composite score, $F(8,149) = 1.07$, $p = .39$.

Major area of study was filtered to include only the majors with the highest number of respondents. Overall, major area of study was not significantly related to the impulse control composite score, $F(5, 83) = .34$, $p = .89$. First language was also not significantly related to the impulse control composite score, $F(6, 151) = .35$, $p = .91$. Lastly, birth country was not significantly related to the impulse control composite score, $F(14, 143) = .31$, $p = .99$.

Trail Making Test A and B (TMT-A and TMT-B)

The relationship between the Trail Making Test A (TMT-A) raw and T scores and the demographic and personal variables outlined in Appendix C and other objective measures (i.e., ImPACT, TOPF, D-KEFS VF) was investigated using Spearman correlations (when correlating with continuous variables) and point-biserial correlations (when correlating with dichotomous variables). TOPF correlations with TMT-A and B raw and T-scores can be found in Table 8. Verbal memory composite, visual memory composite, visual motor speed composite, reaction time composite, and impulse control composite correlations can be found in Table 11. Results of two-tailed bivariate Spearman correlations among TMT-A raw and T scores and demographic and personal variables revealed one significant weak correlation between TMT-A T score and years of education (see Table 13).

Results of two-tailed bivariate Spearman correlations among TMT-B raw and T scores and demographic and personal variables revealed many significant correlations (see Table 13).

Table 13.

TMT-A and B Raw and T Score Correlations with Demographic and Personal Variables

	Years of Education	Age	GPA	Grade 7/8 Math	Grade 7/8 Language	Grade 7/8 Science	Grade 7/8 Social Studies	Grade 7/8 Art	Parent 1/Mother Education	Parent 2/Father Education	Birth Order	SES Ladder	Family Income
TMT-A Raw Score	.116	.017	-.068	.100	.016	.046	-.018	.001	-.079	-.121	-.055	-.141	-.077
TMT-A T	-.207**	-.080	.027	-.087	.006	-.027	.080	-.030	.095	.123	.063	.095	.085
TMT-B Raw Score	.112	.070	-.318**	-.380**	-.186*	-.298**	-.219**	-.110	-.142	-.211**	-.074	-.200*	-.177
TMT-B T	-.159	-.076	.205*	.340**	.096	.256*	.109	-.096	.157	.214*	.073	.120	.232*

Note. *p <.05, **p < .01. A two-tailed bivariate Spearman correlation was conducted to account for non-normal distributions. Significant personal variables for consideration included years of education (more years of education was related to a lower TMT-A T score, despite there being no significant relationship between TMT-A raw score and years of education), GPA, grade 7/8 math, grade 7/8 language, grade 7/8 science, grade 7/8 social studies, parent 2/father's education, and SES ladder (i.e., impression of one's own SES).

Results of correlations between the TMT-A raw and T scores and the remaining objective test scores revealed several significant mild to moderate correlations (see Table 14). Results of correlations between the TMT-B raw and T scores and the remaining objective test scores also revealed several significant correlations (see Table 14).

Table 14.

TMT-A and B Raw and T Score Correlations with Objective Cognitive Test Scores

	Trails B Raw	Trails B T-Score	VF-LF Raw	VF-LF ss	VF-CF Raw	VF-CF ss	VF CS # Raw	VF CS # ss	VF Acc Raw	VF Acc ss
TMT-A Raw	.470**	-.469**	-.101	-.097	-.186*	-.175*	-.093	-.118	-.101	-.094
TMT-A T	-.450**	.498**	.092	.108	.189*	.179*	.067	.095	.059	.064
TMT-B Raw	1.0	-.943**	-.159*	-.180*	-.159*	-.151	-.168*	-.175*	-.125	-.133
TMT-B T	-.943**	1.0	.229	.246*	.206*	.206**	.093	.105	.063	.070

Note. *p <.05, **p < .01. A two-tailed bivariate Spearman correlation was conducted to account for non-normal distributions. Significant correlations to consider appear in bold.

Dichotomous variables were also evaluated, and one significant but weak correlation was identified between TMT-A raw score and sex (see Table 15). This suggests that females performed better in terms of TMT-A raw score. One significant but weak correlation was identified between TMT-B T score and handedness (see Table 15). This relationship suggests that a higher TMT-B T score was weakly related to left-handedness.

Table 15.

TMT-A and B Raw and T Score Correlations with Dichotomous Variables

	Sex	Handedness	Full/Part Time Studies	Scholarship (Yes/No)	SLD Read	SLD Write	SLD Math	ASD	SLI	Speech Disorder	ADHD	Other
TMT-A Raw	**.161***	-.133	-.116	-.021	-.050	-.072	.105	.034	-.004	-.004	-.032	-.083
TMT-A T	-.151	.115	.100	.235	.071	.082	.130	-.034	-.035	-.035	.023	.084
TMT-B Raw	.156	-.147	-.001	-.083	-.038	-.037	-.040	-.063	-.050	-.050	.060	.090
TMT-B T	-.161	**.199***	-.019	-0.019	.075	.009	.025	.020	.047	.047	-.102	-.051

Note. *p <.05, **p < .01 A point-biserial correlation was conducted using bootstrapping to account for non-normal distributions. Significant correlations for consideration included sex and handedness.

One-way ANOVA analyses were conducted when evaluating the mean difference between TMT-A raw and T score and categorical demographic and personal variables. Results did not change when analyses were conducted with a filter to only look at conditions with the most cases, except where indicated. Race was not significantly related to TMT-A raw score, $F(5, 152) = .79$, $p = .56$, or TMT-A T score, $F(5, 152) = .33$, $p = .90$. Sport was also not significantly related to TMT-A raw score differences, $F(8, 149) = .84$, $p = .57$, or TMT-A T score, $F(8, 149) = .76$, $p = .64$.

Major area of study was also not significantly related to TMT-A raw score, F(42, 115) = 1.39, $p = .09$, or TMT-A T score, F(42, 115) = .99, $p = .50$. First language was filtered to include only the previously mentioned most common languages. TMT-A raw score was not significantly related to first language, $F(1, 141) = 2.46$, $p = .12$, nor was

TMT-A T score, $F(1, 141) = .49$, $p = .48$. Lastly, birth country was significantly related to TMT-A raw score, $F(14, 143) = 1.84$, $p = .04$, but not to TMT-A T score, $F(14, 143) = 1.12$, $p = .35$. Post-hoc tests could not be completed due to several groups containing only one observation. When filtered to include only Canada and the USA as done previously, there was no significant difference in TMT-A raw score between groups, $F(1, 136) = .51$, $p = .48$, and in TMT-A T scores between groups, $F(1, 136) = .02$, $p = .90$.

There were no racial differences for TMT-B T score, $F(5, 151) = .82$, $p = .54$. When a filter was applied to include only the Racial categories with the highest number of respondents, there was still heterogeneity of variance for TMT-B raw, and no significant difference between them for the TMT-B T score, $F(1, 134) = 1.19$, $p = .66$. Heterogeneity of variance was also present for sport and TMT-B raw score but not TMT-B T score, where no significant relationship was found, $F(8, 148) = 1.61$, $p = .13$. When the two sports with the smallest sample size were removed, there was still heterogeneity of variance for the TMT-B raw score, and no significant difference for TMT-B T score, $F(5, 141) = 1.38$, $p = .24$.

There was no significant difference in TMT-B T scores across major areas of study, $F(5, 82) = .34$, $p = .89$. First language was filtered to include only the previously mentioned most common languages. TMT-B raw score still failed Levene's statistic ($p < .05$) and TMT-B T score was not significantly related to first language, $F(1, 140) = .02$, $p = .89$. Lastly, TMT-B raw score was again heterogeneous in terms of variance for birth country (Levene's statistic, $p < .05$) and TMT-B T score was not significantly different across groups, $F(14, 142) = .85$, $p = .62$.

Delis-Kaplan Executive Function System Verbal Fluency (D-KEFS VF)

The relationship between the D-KEFS verbal fluency letter fluency (D-KEFS VF-LF), D-KEFS verbal fluency category fluency (D-KEFS CF), D-KEFS verbal fluency category switching total correct (D-KEFS CS #), and D-KEFS verbal fluency category switching accuracy (D-KEFS CS Acc) raw and scaled scores and the demographic and personal variables outlined in Appendix C and other objective measures (i.e., ImPACT, TOPF, TMT-A and B) was investigated using Spearman correlations (when correlating with continuous variables) and point-biserial correlations (when correlating with dichotomous variables). TOPF correlations with D-KEFS verbal fluency raw and scaled scores for all conditions can be found in Table 8. Verbal memory composite, visual memory composite, visual motor speed composite, reaction time composite, and impulse control composite correlations can be found in Table 11. TMT-A and B correlations can be found in Table 13.

Results of two-tailed bivariate Spearman correlations among D-KEFS VF-LF raw and scaled scores and demographic and personal variables revealed one significant but weak positive correlation between D-KEFS VF-LF raw score and years of education. Correlations among D-KEFS VF-CF raw and scaled scores and demographic and personal variables revealed no significant correlations. Correlations among D-KEFS VF-CS number correct and accuracy raw and scaled scores and demographic and personal variables revealed several significant but weak correlations (see Table 16).

Table 16.

Verbal Fluency Raw and Scaled Score Correlations with Demographic and Personal Variables

	Years of Education	Age	GPA	Grade 7/8 Math	Grade 7/8 Language	Grade 7/8 Science	Grade 7/8 Social Studies	Grade 7/8 Art	Parent 1/Mother Education	Parent 2/Father Education	Birth Order	SES Ladder	Family Income
VF-LF Raw	.165*	.117	-.021	.006	.004	.027	.027	.128	.037	.081	-.023	.122	-.018
VF-LF ss	.080	-.023	-.002	-.050	-.044	-.020	-.020	.099	.052	.098	-.028	.101	-.010
VF-CF Raw	.013	-.090	.014	-.081	-.031	.000	-.071	.093	.066	.149	.028	.091	.013
VF-CF ss	-.014	-.070	.003	-.084	-.004	.008	-.048	.121	.109	.144	.049	.087	.055
VF-CS # Raw	.144	.089	.195*	-.102	-.137	-.132	-.226**	-.017	-.027	.053	.031	.003	-.062
VF-CS # ss	.045	.008	.186*	-.101	-.125	-.118	-.212**	.000	.011	.021	.085	-.002	-.015
VF-CS Acc Raw	.143	.085	.184	-.103	-.139	-.132	-.239**	-.052	-.030	.005	.079	-.014	-.048
VF-CS Acc ss	.014	-.054	.216*	-.124	-.154	-.145	-.244**	-.060	.032	.032	-.014	.126	-.025

Note. *p <.05, **p < .01. A two-tailed bivariate Spearman correlation was conducted to account for non-normal distributions. Significant variables for consideration included years of education, GPA, and grade 7/8 social studies grades.

Results of correlations between the D-KEFS VF raw and scaled scores and the remaining objective test scores (i.e., the other D-KEFS VF conditions) revealed many significant correlations (see Table 17). Results of correlations between the D-KEFS VF-CF raw and scaled scores and the remaining objective test scores (i.e., the remaining D-KEFS VF conditions), as well as D-KEFS VF-CS raw and scaled scores revealed several significant correlations (see Table 17).

Table 17.

Verbal Fluency Raw and Scaled Score Correlations with Remaining Objective Cognitive Test Scores

	VF-CF Raw	VF-CF ss	VF CS # Raw	VF CS # ss	VF CS Acc Raw	VF CS Acc ss
VF-LF Raw	.517**	.506**	.303**	.286**	.246**	.233**
VF-LF ss	.518**	.493**	.297**	.268**	.242**	.228**
VF-CF Raw	1.0		.510**	.509**	.465**	.468**
VF-CF ss		1.0	.461**	.478**	.430**	.454**

Note. *p <.05, **p < .01. A two-tailed bivariate Spearman correlation was conducted to account for non-normal distributions. All correlations were significant.

Dichotomous variables were also evaluated, and no significant correlations were found for D-KEFS VF-LF raw or scaled scores. No significant correlations were found for D-KEFS VF-CF raw or scaled scores. Significant correlations were found between VF-CS number correct raw score and sex and VF-CS number correct scaled score and several variables. Significant correlations were also found between VF-CS accuracy raw score and several variables (see Table 18).

Table 18.

D-KEFS Verbal Fluency Raw and Scaled Score Correlations with Dichotomous Variables

	Sex	Handedness	Full/Part Time Studies	Scholarship (Yes/No)	SLD Read	SLD Write	SLD Math	ASD	SLI	Speech Disorder	ADHD	Other
D-KEFS VF-LF raw	.004	.029	.068	-.058	.082	-.078	.084	-.026	.014	.014	.096	.095
D-KEFS VF-LF ss	-.042	.027	.051	-.100	.091	-.066	.065	-.042	.005	.005	.108	.109
D-KEFS VF-CF raw	-.004	.028	-.022	.009	-.045	-.040	-.024	-.085	.022	.022	.035	.126
D-KEFS VF-CF ss	.022	.000	-.055	.016	-.084	-.100	-.041	-.100	.017	.017	.102	.113
D-KEFS VF-CS # raw	**-.197***	.066	.113	.125	.067	-.023	.097	.085	.084	.084	.086	.137
D-KEFS VF-CS # ss	**-.210****	.040	**.169***	**.199***	.054	-.018	.078	.068	.067	.067	.130	.129
D-KEFS VF-CS Acc raw	**-.226****	.048	.117	.124	.031	-.049	.047	.075	.075	.075	.103	**.157***
D-KEFS VF-CS Acc ss	**-.231***	.048	.117	.124	.031	-.049	.047	.075	.075	.075	.103	**.157***

Note. *p <.05, **p < .01 A point-biserial correlation was conducted using bootstrapping to account for non-normal distributions. Significant variables for consideration included sex, full vs. part-time studies, scholarship, and other disorder diagnosis.

One-way ANOVA analyses were conducted when evaluating the mean difference between D-KEFS VF-LF raw and scaled scores and categorical demographic and personal variables. Results did not change when analyses were conducted with a filter to only look at conditions with the most cases, except where indicated. There were no racial differences for VF-LF raw score, F(5, 151) = .80, p = .55, or VF-LF scaled score, F(5, 151) = .43, p = .83. Sport was not significantly related to VF-LF raw score, $F(8, 148)$ = .99, p = .45, or scaled score, $F(8, 148) = 1.10$, p = .37.

There was no significant difference across major areas of study for VF-LF raw score, $F(42, 114) = .97$, p = .53, or VF-LF scaled score, $F(42, 114) = 1.06$, p = .40. There was no significant difference between first language spoken and VF-LF raw score, $F(6, 150) = 1.78$, p = .11, or VF-LF scaled score, $F(6, 150) = 2.02$, p = .07. Lastly, VF-LF raw score was not significantly different across groups for birth country, $F(14, 142) = 1.34$, p = .18, nor was VF-LF scaled score, $F(14, 142) = 1.34$, p = .19.

There were no racial differences for VF-CF raw score, $F(5, 151) = .09$, p = .99, and heterogeneity of variance existed for VF-CF scaled score (Levene's statistic, $p > .05$). Sport was not significantly related to VF-CF raw score, $F(8, 148) = 1.46$, p = .18, or scaled score, $F(8, 148) = 1.55$, p = .15.

There was no significant difference across major areas of study for VF-CF raw score, $F(42, 114)$ = .1,13, p = .30, or VF-CF scaled score, $F(42, 114) = .93$, p = .60. There was a significant difference between first language spoken and VF-CF raw score, $F(6, 150) = 3.40$, p = .004, and for VF-CF scaled score, $F(6, 150) = 4.36$, $p < .001$. First language was filtered to remove English+Arabic due to small sample size (n=1). and there was still a significant difference for VF-CF raw score, $F(3, 145) = 3.12$, p = .03, and VF-CF scaled score, $F(3, 145) = 3.93$, p = .01. The Tukey-Kramer post-hoc test was used due to unequal sample sizes across groups and revealed a significant difference only between English and English+Other language for VF-CF scaled scores, whereby English speakers performed significantly better (M = 13.64, SD = 3.30) than those who spoke both English and another language (*M* = 9.25, *SD* = 3.10). Lastly, VF-CF raw score was not significantly different across groups for birth country, $F(14, 142) = 1.01$, $p = .45$, nor was VF-CF scaled score, $F(14, 142) = 1.27$, $p = .23$.

One-way ANOVA analyses were conducted when evaluating mean difference between D-KEFS VF-CS raw and scaled scores for both the number of correct responses and switching accuracy with categorical demographic and personal variables. There were no racial differences for VF-CS raw scores for number correct, $F(5, 151) = 1.22$, $p = .30$, switching accuracy, $F(5, 151) = .77$, $p = .58$, VF-CS scaled score for number correct, $F(5, 151) = 1.35$, $p = .25$, or switching accuracy, $F(5, 151) = .88$, $p = .50$. When a filter was applied to include only the race categories with the highest number of respondents, there was a significant difference for the VF-CS number correct raw score, $F(1, 133) = 4.88$, $p = .03$, but not for VF-CS switching accuracy raw score, $F(1, 133) = 2.31$, $p = .13$. White athletes produced more correct responses on this condition (*M* = 14.47, *SD* = 2.58) than Black athletes (*M* = 13.31, *SD* = 2.90), however sample sizes are unequal (n=100 vs. n=35). VF-CS number correct scaled score was also significant, $F(1, 133) = 6.17$, $p = .01$, but VF-CS switching accuracy scaled score was not, $F(1, 133) = 3.13$, $p = .08$. White athletes' scaled scores on this condition (*M* = 11.63, *SD* = 3.38) were higher than Black athletes (*M* = 9.97, *SD* = 3.47).

Sport was not significantly related to VF-CS number correct raw score, $F(8, 148) = 1.64$, $p = .12$, but it was significantly related to VF-CS switching accuracy raw score, $F(8, 148) = 2.12$, $p = .04$. VF-CS number correct scaled score was also not significant, $F(8, 148) = 1.82$, $p = .08$, but VF-CS switching accuracy scaled score was significant, $F(8, 148) = 2.15$, $p = .03$. When the two sports with the smallest sample size were removed, there was no significant difference across groups for VF-CS number correct raw score, $F(5, 141) = 1.54$, $p = .18$, but VF-CS switching accuracy raw score was significant, $F(5, 141).= 2.42$, $p = .04$. The Tukey-Kramer post-hoc test was used to investigate this further but revealed no significant comparisons between sports. VF-CS number correct scaled score was not significant, $F(5, 141) = 1.76$, $p = .13$, nor was VF-CS switching accuracy scaled score, F $(5, 141) = 2.23$, $p = .05$.

There was no significant difference across major areas of study for VF-CS number correct raw score, $F(42, 114) = .61$, $p = .97$, and VF-CS switching accuracy raw score, $F(42, 114) = .64$, $p = .95$. VF-CS number correct scaled score was also not significant, $F(42, 114) = .75$, $p = .85$, nor was VF-CS switching accuracy scaled score, $F(42, 114) = .65$, $p = .95$.

There was no significant difference between first language spoken and VF-CS number correct raw score, $F(6, 150) = 1.18$, $p = .32$, or VF-CS switching accuracy raw score, F(6, 150) = .96, $p = .46$. VF-CS number correct scaled score was not significant, $F(6, 150) = 1.24$, $p = .29$, nor was VF-CS switching accuracy scaled score, $F(6, 150) = 1.10$, $p = .36$. First language was filtered to include the two languages with the greatest sample sizes, and there was a significant difference for VF-CS number correct raw score, F(1, 140) = 5.91, $p = .02$, such that English athletes ($M = 14.42$, $SD = 2.60$) produced a higher number of correct responses than French speaking athletes ($M = 12.22$, $SD = 2.67$). It is important to remember, however, that most athletes were English speakers, and there were only 10 French speakers (see Table 4). VF-CS switching accuracy raw score was not significant, $F(1, 140) = 3.47$, $p = .07$. VF-CS number correct scaled score was also significant, $F(1, 140) = 6.37$, $p = .01$, such that English athletes had a higher scaled score (*M* = 11.46, *SD* = 3.33) than French speaking athletes (*M* = 8.56, *SD* = 3.58). VF-CS switching accuracy scaled score was also significant, $F(1, 140) = 4.47$, $p = .04$, such that English athletes (*M* = 11.93, *SD* = 9.78) had a higher switching accuracy score than French speaking athletes (*M* = 9.78, *SD* = 2.22).

Lastly, VF-CS number correct raw score was not significantly different across groups for birth country, $F(14, 142) = 1.49$, $p = .12$, nor was VF-CS switching accuracy raw score, $F(14, 142) = 1.26$, $p = .24$. VF-CS number correct scaled score was also not significant, $F(14, 142) = 1.38$, $p = .17$, nor was VF-CS switching accuracy scaled score, $F(14, 142) = 1.15$, $p = .32$.

Discussion

The current study aimed to investigate the relationships between demographic and personal variables and cognitive test scores at baseline to determine the most relevant predictors of premorbid cognitive functioning in a varsity athletics setting. The sport neuropsychology world has grappled with issues surrounding baseline testing for years, and better understanding the variables that contribute to premorbid functioning estimates provides an important first step towards improving return-to-play decision making. The utility of the comparison between post-injury and baseline test data in return-to-play decision making is based on the integrity of the baseline data, and there is insufficient evidence to suggest that having baseline test results to compare to is superior to not having baseline test results in the sporting world (Iverson & Schatz, 2014). If baseline testing is to be shortened or eliminated, methods for estimating premorbid functioning must be optimized.

Across the cognitive tests included in this test battery, several significant correlations were found, ranging from weak ($r = \pm 0.01 - \pm 0.3$) to strong ($r > 0.6$). Regarding the Test of Premorbid Functioning (TOPF), all significant relationships were weak. Notably, there was a weak relation between TOPF scores and father/second parent's education (such that better performance on the TOPF was related to the athlete's father's years of education). Past research has indicated that a father's education can be positively related to a child's IQ (Kendler et al., 2015; Neisser et al., 1996), such that higher educational attainment in fathers is often associated with a more stimulating home environment, access to educational resources, and positive attitudes towards learning. Verbal fluency - letter fluency and category switching conditions were also related to TOPF score, which is in keeping with hypotheses about this measure as potentially contributing to our understanding of best estimates of premorbid functioning. Additionally, grade 7/8 self-reported grades were identified as having a relationship to TOPF and several other objective test scores. This may be related to the findings of past research which has shown that IQ scores are correlated with school performance, including grades in middle school (Mackintosh, 2011). Grades 7 and 8 are considered to be a period of relative stability, as research suggests that in high school, the impact of increased stress, anxiety, and depression typically becomes more significant and thus impacts grades (APA, 2020). See Table 19 for a summary of the variables that were significantly related ($p < .05$) to TOPF raw score and Standard Score.

Table 19.

Summary of Significant Findings for the Test of Premorbid Functioning (TOPF)

Test/Score	Significant Correlations	Significant ANOVAs

TOPF Raw Score	• Grade 7/8 Math grades (r = .16) • Father/second parent's Education (r = .17) • VF-LF raw score (r = .16) • VF-LF scaled score (r = .16) • VF-CS number correct raw score (r = .16) • VF-CS number correct scaled score (r = .16) • Full vs. part-time studies (r = .16) • SLD in written expression (r = -.19)	None.
TOPF Standard Score	• Grade 7/8 Math grades (r = .21) • Grade 7/8 social studies grades (r = .17) • Father/second parent's education (r = .18) • VF-LF scaled score (r = .16) • VF-CS number correct scaled score (r = .16) • SLD in written expression (r = -.16)	None.

Note. r values between ± 0 and ± 0.3 are considered to be weak, values between ±0.31 and ±0.6 are considered to be moderate, and values between ±0.61 and ±0.9 indicate strong correlations. Any value equal to ±1.0 indicates a perfect correlation.

The next objective test evaluated was the ImPACT, and this exploration included all its composite scores: verbal memory, visual memory, visual motor speed, and reaction time. Overall, significant relationships ranged from weak to moderate, with moderate correlations occurring between ImPACT composites (e.g., verbal memory and visual memory) and other cognitive test scores (e.g., reaction time and trails A/B scores). Several personal variables were significantly but weakly correlated with the ImPACT composite scores, including grade 7/8 self-reported grades in science, math and language, current GPA, scholarship possession, sex, and sport, with women's hockey players performing better on visual motor speed and reaction time composites than men's football. Notably, better scores on the reaction time composite are represented by lower scores, and a lower score on the impulse control composite is indicative of less errors made throughout the ImPACT test battery. See Table 20 for a summary of the significant relationships found for the ImPACT composite scores.

Table 20.

Summary of Significant Findings for the ImPACT

Test/Score	Significant Correlations	Significant ANOVAs
ImPACT Verbal Memory Composite	• Grade 7/8 science (*r* = .25) • Visual Memory Composite (*r* = .47) • Visual Motor Speed Composite (*r* = .19) • Impulse Control Composite (*r* = -.23)	None.
ImPACT Visual Memory Composite	• Grade 7/8 Math grades (*r* = .21) • Grade 7/8 Science grades (*r* = .20) • Verbal Memory Composite (*r* = .47) • Visual Motor Speed Composite (*r* = .18) • Reaction Time Composite (*r* = -.16) • Impulse Control Composite (*r* = -.18) • Trails B raw score (*r* = -.21) • Trails B T score (*r* = .24) • Possession of a scholarship (*r* = .18)	None.
ImPACT Visual Motor Speed Composite	• GPA (*r* = .21) • Grade 7/8 Math grades (*r* = .27) • Grade 7/8 Language grades (*r* = .17) • Grade 7/8 Science grades (*r* = .23)	• Sport (women's hockey > men's football).

	• Grade 7/8 social studies grades (*r* = .20) • Verbal Memory Composite (*r* = .19) • Visual Memory Composite (*r* = .18) • Reaction Time Composite (*r* = -.55) • Trails A raw score (*r* = -.36) • Trails A T score (*r* = .38) • Trails B raw score (*r* = -.39) • Trails B T score (*r* = .38) • VF-LF raw score (*r* = .29) • VF-LF scaled score (*r* = .30) • VF-CF raw score (*r* = .35) • VF-CF scaled score (*r* = .36) • VF-CS # raw score (*r* = .31) • VF-CS # scaled score (*r* = .26) • VF-CS accuracy raw score (*r* = .29) • VF-CS accuracy scaled score (*r* = .25)	
ImPACT Reaction Time Composite	• Sex (*r* = -.21) • GPA (*r* = -.30) • Grade 7/8 Math grades (*r* = .22) • Grade 7/8 Language grades (*r* = .17) • Grade 7/8 Science grades (*r* = .16) • Visual Memory Composite (*r* = -.16) • Visual Motor Speed Composite (*r* = -.55) • Trails A raw score (*r* = .34) • Trails A T score (*r* = -.25) • Trails B raw score (*r* = .35) • Trails B T score (*r* = -.31) • VF-LF raw score (*r* = -.22) • VF-LF scaled score (*r* = -.23) • VF-CF raw score (*r* = -.34)	• Sport (women's hockey > men's football)

	• VF-CF scaled score (r = -.34) • VF-CS number correct raw score (r = -.26) • VF-CS number correct scaled score (r = -.24) • VF-CS accuracy raw score (r = -.29) • VF-CS accuracy scaled score (r = -.28)	
ImPACT Impulse Control Composite	• Sex (r = .18) • Verbal Memory Composite (r = -.23) • Visual Memory Composite (r = -.18)	None.

Note. r values between ± 0 and ± 0.3 are considered to be weak, values between ±0.31 and ±0.6 are considered to be moderate, and values between ±0.61 and ±0.9 indicate strong correlations. Any value equal to ±1.0 indicates a perfect correlation.

The Trail Making Test (TMT) A and B were also evaluated in terms of their relationships to the personal and demographic variables and other objective test scores. Overall, relationships ranged from weak to moderate, with the moderate relationships including the TMTs correlations with other objective test scores and with GPA. Other variables with weak but significant correlations include sex, years of education, grade 7/8 self-reported grades in math, language, and science, athlete impression of their own socioeconomic status (SES ladder), handedness, and mother/first parent's education. A summary of the significant relationships found for the TMT A and B are found in Table 21.

Table 21.

Summary of Significant Findings for the Trail Making Test A and B

Test/Score	Significant Correlations	Significant ANOVAs
Trail Making Test A Raw Score	• Visual Motor Speed Composite Score (r = -.36)	None.

	• Reaction Time Composite Score (*r* = .34) • Trails B raw score (*r* = .47) • Trails B T-score (*r* = -.47) • VF-CF raw score (*r* = -.19) • VF-CF scaled score (*r* = -.18) • Sex (*r* = .16)	
Trail Making Test A T Score	• Years of Education (*r* = -.21) • Visual Motor Speed Composite Score (*r* = .35) • Reaction Time Composite Score (*r* = -.34) • Trails B raw score (*r* = -.45) • Trails B T score (*r* = .50) • VF-CF raw score (*r* = .19) • VF-CF scaled score (*r* = .18)	None.
Trail Making Test B Raw Score	• GPA (*r* = -.32) • Grade 7/8 Math grades (*r* = .38) • Grade 7/8 Language grades (*r* = .19) • Grade 7/8 Science grades (*r* = .30) • Grade 7/8 Social Studies grades (*r* = .22) • Father/second parent's education (*r* = -.21) • SES ladder (*r* = -.20) • Visual Memory Composite Score (*r* = -.21) • Visual Motor Speed Composite Score (*r* = -.39) • Reaction Time Composite Score (*r* = .35) • Trails A raw score (*r* = .47) • Trails A T score (*r* = -.45) • VF-LF raw score (*r* = -.16) • VF-LF scaled score (*r* = -.18) • VF-CF raw score (*r* = -.16) • VF-CF scaled score (*r* = -.15) • VF-CS number correct raw score (*r* = -.17) • VF-CS number correct scaled score (*r* = -.18)	None.
Trail Making Test B T Score	• Years of education (*r* = -.16) • GPA (*r* = .21) • Grade 7/8 Math grades (*r* = .34)	None.

	• Grade 7/8 Science grades (*r* = .26) • Father/second parent's education (*r* = 21) • Visual Memory Composite Score (*r* = .16) • Visual Motor Speed Composite Score (*r* = .37) • Reaction Time Composite Score (*r* = -.27) • Trails A raw score (*r* = -.47) • Trails A T score (*r* = .48) • VF-LF scaled score (*r* = .19) • VF-CF raw score (*r* = .19) • VF-CF scaled score (*r* = .20) • Handedness (*r* = .20)

Note. *r* values between ± 0 and ± 0.3 are considered to be weak, values between ±0.31 and ±0.6 are considered to be moderate, and values between ±0.61 and ±0.9 indicate strong correlations. Any value equal to ±1.0 indicates a perfect correlation.

The D-KEFS Verbal Fluency subtests were also evaluated in terms of their relationships to the personal and demographic variables and other objective test scores. Significant relationships ranged from weak to strong, with strong correlations existing between the two scores provided for verbal fluency category switching (i.e., number correct and accuracy), and moderate correlations existing between verbal fluency scores and other objective cognitive test scores. Significant but weak correlations existed between the D-KEFS verbal fluency scores and years of education, GPA, sex, full/part time studies, and scholarship. There was a significant difference in D-KEFS verbal fluency category switching based on first language, where English speakers tended to perform better than French speakers. There was also a significant difference in D-KEFS verbal fluency category switching number correct based on race, whereby White athletes tended to perform better than Black athletes, and first language, whereby English speakers tended to perform better than French speakers. It is possible that this could be due to bilingual disadvantages in verbal fluency, such that there is cross-language interference between the two languages (Sandoval et al., 2010). See Table 22 for a summary of all significant relationships found for the D-KEFS Verbal Fluency subtests.

Table 22.

Summary of Significant Findings for D-KEFS Verbal Fluency

Test/Score	Significant Correlations	Significant ANOVAs
D-KEFS VF-LF Raw Score	• Years of education (*r* = .17) • TOPF raw score (*r* = .16) • Visual Motor Speed Composite (*r* = .29) • Reaction Time Composite (*r* = -.22) • Trails B raw score (*r* = -.16) • VF-CF raw score (*r* = .52) • VF-CF scaled score (*r* = .51) • VF-CS number correct raw score (*r* = .30) • VF-CS number correct scaled score (*r* = .29) • VF-CS accuracy raw score (*r* = .25) • VF-CS accuracy scaled score (*r* = .23)	None.
D-KEFS VF-LF Scaled Score	• TOPF raw score (*r* = .16) • TOPF Standard Score (*r* = .16) • Visual Motor Speed Composite (*r* = .30) • Reaction Time Composite (*r* = -.23) • Trails B raw score (*r* = -.18) • Trails B T score (*r* = .17) • VF-CF raw score (*r* = .52) • VF-CF scaled score (*r* = .49) • VF-CS number correct raw score (*r* = .30) • VF-CS number correct scaled score (*r* = .27) • VF-CS accuracy raw score (*r* = .24) • VF-CS accuracy scaled score (*r* = .23)	None.

D-KEFS VF-CF Raw Score	• Visual Motor Speed Composite (*r* = .35) • Trails A raw score (*r* = -.19) • Trails A T score (*r* = .20) • Trails B raw score (*r* = -.16) • Trails B T score (*r* = .20) • VF-CS number correct raw score (*r* = .51) • VF-CS number correct scaled score (*r* = .51) • VF-CS accuracy raw score (*r* = .47) • VF-CS accuracy scaled score (*r* = .47) • Reaction Time composite (*r* = -.34)	• First language (English > English + Other language)
D-KEFS VF-CF Scaled Score	• Visual Motor Speed Composite (*r* = .36) • Reaction Time Composite (*r* = -.34) • Trails A raw score (*r* = -.18) • Trails A T score (*r* = .20) • Trails B raw score (*r* = -.15) • Trails B T score (*r* = .21) • VF-CS number correct raw score (*r* = .46) • VF-CS number correct scaled score (*r* = .48) • VF-CS accuracy raw score (*r* = .43) • VF-CS accuracy scaled score (*r* = .45)	• First language (English > English + Other language)
D-KEFS VF-CS # Correct Raw Score	• GPA (*r* = .20) • Grade 7/8 Social Studies grades (*r* = -.23) • TOPF raw score (*r* = .16) • Visual Motor Speed Composite (*r* = .31) • Reaction Time Composite Score (*r* = -.26) • Trails B raw score (*r* = -.17) • VF-LF raw score (*r* = .30) • VF-LF scaled score (*r* = .30) • VF-CF raw score (*r* = .51) • VF-CF scaled score (*r* = .46) • VF-CS Accuracy raw score (*r* = .92) • VF-CS accuracy scaled score (*r* = .88)	• Race (White > Black) • First language (English > French)

D-KEFS VF-CS # Correct Scaled Score	• Sex (*r* = -.20) • GPA (*r* = .19) • Grade 7/8 Social Studies grades (*r* = -.21) • TOPF raw score (*r* = .16) • TOPF standard score (*r* = .16) • Visual Motor Speed Composite (*r* = .26) • Reaction Time Composite (*r* = -.24) • Trails B raw score (*r* = -.18) • VF-LF raw score (*r* = .29) • VF-LF scaled score (*r* = .27) • VF-CF raw score (*r* = .51) • VF-CF scaled score (*r* = .48) • VF-CS accuracy raw score (*r* = .88) • VF-CS accuracy scaled score (*r* = .91)	• Race (White > Black) • First language (English > French)
D-KEFS VF-CS Accuracy Raw Score	• Sex (*r* = -.21) • Full/Part-time studies (*r* = .17) • Scholarship (*r* = .20) • Grade 7/8 Social Studies grades (*r* = -.24) • Visual Motor Speed Composite (*r* = .29) • Reaction Time Composite (*r* = -.29) • VF-LF raw score (*r* = .25) • VF-LF scaled score (*r* = .24) • VF-CF raw score (*r* = .47) • VF-CF scaled score (*r* = .43) • VF-CS number correct raw score (*r* = .92) • VF-CS number correct scaled score (*r* = .88)	
D-KEFS VF-CS Accuracy Scaled Score	• Sex (*r* = -.23) • GPA (*r* = .22) • Grade 7/8 Social Studies grades (*r* = -.24) • Visual Motor Speed Composite (*r* = .25) • Reaction Time Compsite (*r* = -.28) • VF-LF raw score (*r* = .23) • VF-LF scaled score (*r* = .23) • VF-CF raw score (*r* = -.34) • VF-CF scaled score (*r* = .91)	

	• VF-CS number correct raw score (r = .88) • VF-CS number correct scaled score (r = .91) • Sex (r = -.23)

Note. r values between ± 0 and ± 0.3 are considered to be weak, values between ±0.31 and ±0.6 are considered to be moderate, and values between ±0.61 and ±0.9 indicate strong correlations. Any value equal to ±1.0 indicates a perfect correlation.

As such, this study sought to identify the relationships between a vast range of personal and demographic variables and objective psychological test scores in an exploratory fashion. One of the most salient limitations to this sort of analysis is the fact that a wide range of factors can impact one's cognitive performance, as can be seen by the number of weak to moderate correlations found between personal and demographic variables and cognitive test scores. Additionally, external influences can include physiological factors such as a poor night's sleep, to psychological factors such as significant life stress, depressive symptomatology, or performance-based anxiety. Although this study has attempted to capture many factors beyond what is typical for a baseline assessment, it is difficult to pinpoint factors that directly influence performance confidently. Additionally, some factors such as education and parent education are much more complex than they may appear just based on "years of education." For example, measuring educational quality is not straightforward and potentially encompasses interrelated factors such as achievement, school-specific features, and regional SES.

Another limitation that impacted this study was that although the total sample size was quite large, the sample size within groups, especially minority groups, often contained few athletes, thus muddying the waters when attempting to make overarching conclusions and account for diversity of response. For example, when looking at the impact of first language on test performance, the sample sizes were extremely unequal and included 133 English speakers, 10 French speakers, 1 English and Arabic speaker, 3 English and French speakers, 4 English and other language speakers, and 3 other language speakers. Because of unequal variances, this variable was filtered to include only English and French speakers, and it was determined that English speakers performed better on two D-KEFS verbal fluency subtests than French speakers. Although it makes sense that those who spoke English as their first language rather than French would perform better on an English-based word generation task, this finding comes from the inclusion of only 10 French speakers. Research in this area is mixed, with some arguing that different backgrounds of participant groups in terms of first language and culture do not contribute significantly to variability in performance on this task (Pekkala et al., 2008), and some arguing that compared to monolinguals, bilinguals show a disadvantage in category fluency due to relying on more rapid first language retrieval than monolinguals (Lehtinen et al., 2023). As such, small sample size within groups is a significant limitation and these results should be considered preliminary until future research is able to replicate it in a larger sample. It is also notable that uneven group sizes can create several challenges and considerations. Even if Levene's test determined homogeneity of variance, the statistical power of the ANOVAs conducted were affected, since groups with smaller sample sizes contribute less information to the analysis and possess an unfair representation of that group.

Additionally, this study was designed specifically with a varsity athlete population in mind, and thus was not designed to be generalizable to non-athlete populations as the sample possessed a restricted range in terms of age and years of education and was composed of only varsity athletes. The sample was also primarily

White, born in Canada, English-speaking, and a part of the men's football team. There was also a vast range of major areas of study, with Kinesiology being the most common. A larger and more representative sample size is also needed to be able to generalize to other varsity athletics settings that do not mimic this sample's composition and to draw conclusions about differences in group performance for groups with very few athletes.

In conclusion, the current study was the first to investigate the relationships between a wide range of demographic and personal variables and objective psychological test scores in a varsity athletics setting. Significant relationships were found but were generally weakly correlated or lacked an appropriate sample size. As such, this study gives neuropsychologists an idea of which variables to consider that could possibly be contributing to test performance on the tests included in this battery (variables with significant correlations included: parental education, full vs. part-time studies, possession of a scholarship, grade 7/8 self-reported grades, current GPA, sex, race, sport, years of education, handedness, impression of SES, first language), and acts as a first step towards a "precision medicine" approach to this complicated and important problem in sport neuropsychology. This study also promotes the use of a wider range of variables more broadly, as it demonstrates that test scores are related to numerous different variables, therefore promoting the use of a biopsychosocial approach to neuropsychological assessment.

CHAPTER 4

Study II: Using Verbal Fluency Word Quality as a Predictor Variable of Level of Premorbid Functioning

Tests of verbal fluency are often a component of neuropsychological test batteries because they provide insight into various aspects of cognitive functioning, including language skills, executive functioning, attention, and mental flexibility. Letter fluency tasks are heavily mediated by the left frontal lobe and semantic fluency tasks are mediated by frontal and temporal structures (Henry & Crawford, 2004). Several studies have reported significant correlations between verbal fluency and premorbid intellectual ability as measured by oral word reading (see Crawford et al., 1992, Harnett et al., 2004; Ardila et al. 2000). These correlations are unsurprising since tests of verbal fluency and oral word reading are both dependent on verbal ability, and both tests are closely linked with verbal intelligence (Ardila et al., 2000).

Age of acquisition (AoA) is a psycholinguistic variable that refers to the age at which a word is commonly learned. For example, the word "mom" is typically learned at a younger age than the word "narwhal." It has generally been found that words that are more frequently used (i.e., high frequency words) are learned earlier than others, and that earlier-acquired words are processed more efficiently than later-acquired words (Meschyan et al., 2002; Morrison et al., 1992; Catling & Elsherif, 2020). Therefore, both word frequency and word age of acquisition are word quality measures that play a fundamental role in lexical retrieval (Meschyan & Hernandez, 2002). As such, performance on tasks of verbal fluency in terms of the quality of words produced could provide important information regarding one's level of premorbid functioning.

Several studies have reported significant correlations between verbal fluency and premorbid intellectual ability as measured by oral word reading (Crawfored et al., 1992; Harnett et al., 2004). Because tests of verbal fluency and those often used in the estimation of premorbid ability, including oral word reading tests, are both dependent on verbal ability, it is unsurprising that a relationship has been identified (Ardila et al., 2000). Currently, no known published research has demonstrated the use of indices of the quality of the words produced within the verbal fluency task as a direct measure of premorbid functioning. A poster presentation by Abeare and Seguin (2014) found that those with higher NAART estimated IQ scores were more likely to generate words on a verbal fluency phonemic fluency task that have a higher AoA value, and that maximum AoA produced is a better predictor of estimated IQ than the NAART itself. They also demonstrated that the use of the AoA index and other word quality information could serve as an important predictor variable when developing an estimate of one's premorbid functioning.

Study Aims and Hypothesis

This study aims to replicate, improve, and expand upon Abeare and Seguin (2014)'s methodology using the D-KEFS Verbal Fluency (VF) test (letter fluency subtest) and the TOPF instead of the NAART, as they are both measures with more up-to-date normative samples. Specifically, the quality of the words produced by athletes within the D-KEFS VF letter fluency subtest will be examined, based on the word frequency value of each word, and their relationship with the objective test scores and personal and demographic variables examined in Study 1. Abeare and Seguin (2014) used Age of Acquisition (AoA) values, but AoA values were not available for many of the words produced by athletes in this study, leaving significant amounts of missing data, particularly for more obscure words that would ultimately result in a higher AoA value. Importantly, AoA is related to word frequency (Juhasz, 2005), and both AoA and word frequency reliably facilitate the speed and accuracy of word retrieval (Meschyan & Hernandez, 2002). It has also been suggested that the AoA of words is related to the fact that an earlier learned word has been encountered more often (Ghyselinck et al. 2004), and so the use of word frequency values in this study is justified and important.

Studies in psycholinguistics have generally found that there is a negative correlation between word frequency and age of acquisition, such that words that are learned earlier are used more frequently. For instance, a study by Brysbaert et al. (2014) found a negative correlation (r = -.46) between word frequency and age of acquisition. A study by Gilhooly and Logie (1980) also reported a negative correlation (r = -.72) between a related measure of word familiarity (i.e., self-reported familiarity using a 7-point Likert scale) and age of acquisition. It is important to note, however, that correlation coefficients between word frequency and AoA vary across different datasets, languages, methodologies, and timepoints. However, the negative correlation between these variables is a consistent finding in psycholinguistic research.

The use of word frequency may be a better indicator of AoA, since a larger database of word frequency values is available to be able to assign rankings to a broader range of words than available AoA values. The Corpus of Contemporary American English (COCA; Davies, 2008-) is the most widely used online corpora of English words, with 155 billion English words entered since 2008. It is widely used for research in various fields, including linguistics, psycholinguistics, computational linguistics, sociolinguistics, and beyond. The words contained in this corpus were collected from spoken language, fiction, popular magazines, newspapers, television, blogs, spoken interviews, and academic texts, and contains information used for many literary purposes, including word frequency values that are derived from the frequency of each word's appearance within the COCA. It is current, allowing for a snapshot of the word frequency values of the words provided by athletes in the most up-to-date manner, rather than using AoA database studies which tend to be published 10 or more years ago, and/or contain only hundreds to several thousand AoA word values, resulting in missing data (i.e., Kuperman et al. 2012; Gilhooly, 1980; Morrison & Catriona, 1997; Juhahsz et al., (2015); Scott et al. 2019). Overall, the COCA is a versatile resource that supports a wide range of research endeavors and is useful because of its vast size, diversity of sources, and accessibility. It is continuously updated and maintained to reflect contemporary American English usage. As such, it is hypothesized that much like in Abeare and Seguin's (2014) study, qualities such as word frequency that are related to the AoA of the words produced will be a strong predictor of premorbid functioning and will account for unique variance in the prediction of premorbid functioning. The use of word frequency values derived from the COCA improves upon Abeare and Seguin's (2014) methodology because it allows for a larger database of words with more current linguistic indices than those derived from AoA databases.

Method

Procedure

Following the general procedure described previously, raw athlete data files were used to transcribe the words produced for the three letter conditions of the D-KEFS Verbal Fluency Letter Fluency subtest into an Excel spreadsheet. Each individual word was entered into the Corpus of Contemporary American English (COCA; Davies, 2008-). The word frequency values provided by the COCA were recorded for each word produced across all three trials of the Verbal Fluency – Letter Fluency subtest. These values represent the relative frequency of occurrence of words within the corpus, thus providing information about how often a particular word appears in the corpus relative to other words. For example, the word "ficus" has a word frequency score of 34,377, meaning that it is the 34,377th most frequent word in their database. High frequency words are considered to range from ranking values of 1-5000, medium frequency words range from 5000-25,000, and low frequency words have values greater than 25,000 (Davies, 2008-), with the word "the" being the highest frequency word in their database (i.e., it has a word frequency score of 1). Once each of the athletes' words were assigned a word frequency score across all three D-KEFS VF letter fluency trials, the maximum word frequency score (i.e., least common word), minimum word frequency score (i.e., most common word), and average word frequency score for each athlete across each letter trial was calculated. Minimum/most common word frequency score was included in-keeping with Abeare and Seguin's (2014) methodology which included minimum AoA score, but also to act as an index that gathers information about more common words that were more likely to be used on the verbal fluency task as a strategy to produce words as quickly as possible. Analyses were completed using the maximum, minimum, and average word frequency score for each athlete across all three letter trials combined (trial 1 + trial 2 + trial 3). From this point forward and for ease of understanding, maximum score is also referred to as "least common word score" and minimum score is referred to as "most common word score."

Data Analysis

Preliminary analyses. All analyses were conducted using IBM SPSS Statistics v.29.0.0.0. Prior to conducting the primary analyses, the assumptions of correlation and ANOVA analyses were tested. ANOVA analyses were conducted when evaluating mean differences between nominal/categorical variables and continuous variables. Assumptions of ANOVA include normality of variables (e.g., histograms, Q-Q plot, Kolmogorov-Smirnov test), equal variances of the populations that the samples come from (e.g., boxplots, Bartlett's test), and independence of observations. As mentioned previously, there were significant differences in the number of athletes within several of the groups that were to be evaluated using ANOVA, and thus the variance was not equal. These discrepancies were outlined in Tables 1, 2, 3, 4, and 5, for Study 1. As was done in Study 1, for the variables outlined in Tables 1-5, Levene's test was used to determine homogeneity of variance across each dependent variable, and filtering methods were used to look at only the groups with appropriate sample sizes. Running an ANOVA with a group containing only one observation is not statistically valid, since ANOVA relies upon comparing means across groups, and thus filtering was used to account for this and remove these variables.

Results

The maximum/least common, minimum/most common, and average word frequency rankings obtained for each athlete were correlated with the demographic and personal variables outlined in Appendix C and other objective measures (i.e., TOPF TMT A and B, ImPACT composite scores, D-KEFS VF scores) using Spearman correlations (when correlating with continuous variables) and point-biserial correlations (when correlating with dichotomous variables). See Table 23 for the descriptive statistics for each measure.

Table 23.

Descriptive Statistics for Letter Fluency Maximum, Minimum, and Average Frequency Rankings.

Score	*M* (SD)	Range
Letter Fluency Maximum Word Frequency Score	30561.60 (12260.61)	10630-66523
Letter Fluency Minimum Word Frequency Score	87.76 (135.63)	2-714
Letter Fluency Average Word Frequency Score	5389.13 (1648.18)	2156.71-10586

*Note. N = 145 across all scores.

Results of two-tailed bivariate Spearman correlations among LF word frequency values revealed a significant negative weak correlation between LF maximum/least common word and birth order, $r = -.18$, $p = .04$, and a significant negative weak correlation between LF average and GPA, $r = -.28$, $p = .005$ (see Table 24). This means that those who were born earlier in the birth order (i.e., first born children) tended to produce words with a higher maximum word frequency value (i.e., less common English words) than those born later in the birth order. As well, those with a higher GPA, tended to produce more common words on average.

To normalize the distribution of LF average word frequency and stabilize variance, a natural log transformation was performed. The original variable, LF average word frequency score, was transformed using a common logarithm function (base 10) because there were no zero or negative values in the data. Results of two-tailed bivariate Spearman correlations remained largely consistent with the untransformed data, however grade 7/8 math was now associated with a significant weak correlation, *r* = .18, *p* = .04 (see Table 24).

Table 24.

Letter Fluency Maximum/Least Common, Minimum/Most Common, and Average Word Frequency Score Correlations with Demographic and Personal Variables

	Years of Education	Age	GPA	Grade 7/8 Math	Grade 7/8 Language	Grade 7/8 Science	Grade 7/8 Social Studies	Grade 7/8 Art	Parent 1/Mother Education	Parent 2/Father Education	Birth Order	SES Ladder	Family Income
LF Max/Least Common Score	.055	.053	-.121	.134	.098	.157	.049	.124	-.138	-.058	**-.177***	-.020	-.147
LF Min/Most Common	.029	.039	-.142	.029	.117	-.083	.009	.064	.108	.073	-.058	-.099	-.144
LF Average	-.011	.052	**-.275****	.161	.106	.121	.019	.064	.031	.073	-.023	-.111	-.140
LF Average Log Transformed	.002	.062	**-.283****	**.179***	.103	.120	.021	.130	.007	.068	-.048	-.124	-.111

Note. *p <.05, **p < .01. A two-tailed bivariate Spearman correlation was conducted to account for non-normal distributions. Significant correlations to consider include GPA and birth order.

Regarding objective cognitive testing, only weak significant correlations were found. LF minimum/most common score was negatively correlated with the ImPACT visual memory composite, *r* = -.17, *p* = .04. This means that higher scores on the ImPACT visual memory composite were related to lower LF minimum scores (i.e., the most common words produced). LF maximum/least common score was positively correlated with VF-LF raw score, *r* = .17, *p* = .04, and VF-LF scaled score, *r* = .17, *p* = .04. This means that higher scores on VF-LF were related to higher maximum LF scores (i.e., those that produced more words on VF-LF tended to produce their maximum LF word with a higher value). LF average score was positively correlated with VF-LF raw score, *r* = .19, *p* = .02, VF-LF scaled score, *r* = .18, *p* = .04, and VF-CF scaled score, *r* = .19, *p* = .03. This means that those who produced more words on VF-LF and VF-CF (when compared to the normative sample only; scaled score and not raw score) tended to produce words with a higher LF average score (i.e., less common words). When log transformed, LF average word frequency score correlated only with VF-CF scaled score, $r = .19$, $p = .03$ (see Table 25).

Table 25.

Letter Fluency Maximum/Least Common, Minimum/Most Common, and Average Word Frequency Score Correlations with Objective Test Scores

	ImPACT Verbal Memory Composite	ImPACT Visual Memory Composite	ImPACT VMS Composite	ImPACT Reaction Time Composite	ImPACT IC Composite	TOPF Raw	TOPF SS	Trails A Raw	Trails A T-Score	Trails B Raw	Trails B T-Score	VF-LF Raw	VF-LF ss	VF-CF Raw	VF-CF ss	VF CS # Raw	VF CS # ss	VF Acc Raw	VF Acc ss
LF Max/Least common	-.147	-.055	-.020	-.020	-.088	.135	.144	.116	-.122	.075	-.080	**.172***	**.170***	.121	.123	.022	.010	-.020	-.013
LF Min/most common	.000	**-.168***	-.010	-.007	.090	.079	.085	-.060	.116	-.078	.144	-.060	-.075	.030	.087	-.065	-.035	-.054	-.031
LF Average	-.035	-.057	.030	-.021	--.071	.108	.106	.048	-.047	.019	-.021	**.190***	**.175***	.160	**.187***	.061	.075	.002	.021
LF Average Log Transformed	-.018	-.071	.006	.002	-.099	.142	.139	.078	-.047	.048	-.021	.165	.151	.157	**.186***	.081	.091	.008	.017

Note. *p <.05, **p < .01. A two-tailed bivariate Spearman correlation was conducted to account for non-normal distributions. Significant correlations to consider include ImPACT visual memory, VF-LF raw and ss, and VF-CF ss.

Point-biserial correlations were conducted for dichotomous variables, with no significant correlations between LF maximum, minimum, average, or log transformed average scores and the dichotomous variables (see Table 26).

Table 26.

Letter Fluency Maximum/Least Common, Minimum/Most Common, and Average Word Frequency Score Correlations with Dichotomous Variables

	Sex	Handedness	Full/Part Time Studies	Scholarship (Yes/No)	SLD Read	SLD Write	SLD Math	ASD	ADHD	Other
LF Max/Least Common	.014	.078	-.025	.022	-.100	-.064	.040	.047	-.005	.108
LF Min/Most Common	.019	-.026	-.096	.025	-.083	-.028	-.050	.043	-.004	-.052
LF Average	-.014	.033	-.020	.057	-.146	-.078	.132	.069	-.063	-.013
LF Average Log Transformed	-.013	.022	-.005	.083	-.159	-.084	.124	.075	-.058	-.003

Note. *p <.05. A point-biserial correlation was conducted using bootstrapping to account for non-normal distributions.

One-way ANOVA analyses were conducted when evaluating the relationship (i.e., mean differences) between nominal/categorical variables (independent variable/factor) and the continuous LF variables (dependent variable). The log transformed LF average score was also investigated, but findings were largely consistent with those for the untransformed LF average and are therefore not reported. Levene's test of homogeneity of variables was used for each analysis and is reported if violated. These variables included race, sport, major area of study, previous degree, first language, and birth country. A one-way ANOVA was conducted for race and was not significantly related to LF maximum/least common score, $F(5, 139) = .79$, $p = .56$, LF minimum/most common score, $F(5, 139) = .86$, $p =.51$, or average score, $F(5, 139) = 1.67$, $p = .15$. When filtered to include only the races with the highest number of respondents, as demonstrated in Study 1, no significant differences were found for LF maximum/least common score, $F(1, 122) = 1.42$, $p = .24$ or minimum/most common score, $F(1, 122) = 2.02$, $p = .16$. There was a significant difference found for LF average score, $F(1, 122) = 6.04$, $p = .02$, such that Black athletes tended to produce words with higher frequency scores (i.e., less common words) on average ($M = 6034.28$, $SD = 1612.09$) than White athletes ($M = 5184.47$, $SD = 1708.71$).

Based on a one-way ANOVA, sport was not significantly related to LF maximum/least common score, $F(8, 136) = .14$, $p = .99$, minimum/most common score, $F(8, 136) = .58$, $p = .78$, or average score, $F(8, 137) = .60$, $p = .78$. A filter was applied to remove those who played men's hockey (n=2) and men's soccer (n=1) due to small sample sizes, and sport was still not significantly related to LF maximum/least common score, $F(5, 130) = .37$, $p = .87$, minimum/most common score, $F(5, 130) = .56$, p = .73, or average score, $F(5, 130) = .56$, $p = .73$.

Major area of study was not significantly related to LF maximum/least common score, $F(38, 106) = 1.05$, $p = .41$, but it was significantly related to minimum/most common score $F(38, 106) = 1.60$, $p = .03$. Major area of study was not related to LF average score, $F(38, 106) = 1.18$, $p = .26$. Notably, Levene's statistic was significant, which represents heterogeneity of variance across all three measures reported above. As such, a filter was applied to investigate the relationship between the major areas of study that were most common among athletes, including Kinesiology (n=31), Business Administration (n=25), Criminology (n=12), Sport Management and Leadership (n=11), Psychology (n=9), and Nursing (n=5). Major area of study including these majors only no longer violated the assumption of homogeneity of variances, but was not significantly related to maximum/least common score, $F(5, 77) = .89$, $p = .49$, minimum/most common score, $F(5, 77) = .78$, $p = .57$, or average score, $F(5, 77) = .69$, $p = .64$.

Again, despite sample sizes across groups being small and unequal, homogeneity of variance existed across the first language variable with regards to LF scores ($p > .05$). As such, first language did not significantly relate to LF maximum/least common score, $F(6,138) = .32$, $p = .93$, nor did it relate to minimum/most common score, $F(6, 138) = 2.02$, $p = .07$, or average score, $F(6, 138) = .26$, $p = .96$. A filter was applied to include only the first languages with the most athletes, namely English (n=133) and French (n=10). Again, first language did not significantly relate to LF maximum/least common score, $F(1, 129) = 50$, $p = .48$, minimum/most common score, $F(1, 129) = .07$, $p = .79$, or average score, $F(1, 129) = .001$, $p = .98$.

Birth country did not significantly relate to LF maximum/least common score, $F(13, 131) = .41$, $p = .96$, minimum/most common score, $F(13, 131) = 1.78$, $p = .05$, or average score, $F(13, 131) = .58$, $p = .87$, but Levene's statistic ($p < .05$) suggested heterogeneity of variance existed for the maximum/least common and minimum/most common score. A filter was applied to investigate the birth countries with the most athletes, namely Canada (n=127) and the USA (n=11), and homogeneity of variance was satisfied. Birth country (including only Canada and the USA) did not significantly relate to LF maximum/least common score, $F(1, 124) = .75$, $p = .39$, minimum/most common score, $F(1, 124) = 2.65$, $p = .11$, or average score, $F(1, 124) = 2.56$, $p = .11$.

Discussion

The current study aimed to replicate, improve, and expand upon Abeare and Seguin's (2014) study, whereby they found that the quality of words produced during a verbal fluency task was related to NAART estimated IQ scores. As such, this study used the D-KEFS verbal fluency test's letter fluency (D-KEFS VF LF) sub-task, which is essentially the same task used by Abeare and Seguin (2014), with a more updated normative sample. This study also used the TOPF instead of the NAART, as the TOPF is a more current version of a word-reading task (i.e., the NAART was published in 1989, the TOPF was published in 2009). Specifically, the quality of words produced by athletes within the D-KEFS VF LF subtest was examined based on the word frequency value of each word, and their relationship with the objective test scores, personal, and demographic variables examined in Study 1.

Originally, the purpose of this study was to use Age of Acquisition (AoA) as a measure of word quality produced, just as Abeare and Seguin (2014) did. However, it was found that many of the words produced by athletes were not included in AoA normative databases which resulted in a sample that lacked a representative measure of word quality, particularly for less common words. Additionally, AoA has limitations including variability in the age at which individuals acquire words, subjectivity in determining the exact age the word was acquired, and a general lack of precision since AoA is often measured in broad categories (e.g., early childhood, adolescence, adulthood). Another salient limitation was that the normative databases which included specific AoA values were either published 10+ years ago, or only included several thousand words (i.e., Kuperman et al. 2012; Gilhooly, 1980; Morrison & Catriona, 1997; Juhahsz et al., (2015); Scott et al. 2019).

It was therefore decided that the use of word frequency could be a better indicator of word quality, that is ultimately related to AoA (Juhasz, 2005). Word frequency is another index of word quality that has been determined to reliably facilitate the speed and accuracy of word retrieval (Meschyan & Hernandez, 2002). As well, it was determined that the use of a larger and widely used database, The Corpus of Contemporary American English (COCA; Davies, 2008-) resulted in word frequency values for virtually all of the words that athletes produced on this task. Most importantly, it is current to the words produced by athletes at the time that data collection for this study was completed. Therefore, word frequency is considered to be a more accurate representation of word quality to support the purpose of this study, and to improve upon Abeare and Seguin (2014)'s methodology.

In an effort to add to our understanding of a precision approach to the estimation of premorbid functioning, each athlete's maximum word frequency value (i.e., the highest word frequency value derived from the COCA; the least common word), minimum word frequency value (i.e., the lowest word frequency value derived from the COCA; the most common word), and average word frequency value (i.e., the average of each athlete's word frequency values derived from the COCA) were calculated using all three trials of the D-KEFS VF LF subtest. In the same manner described in Study 1, these values were correlated and compared to the vast range of objective test scores and personal and demographic variables.

LF Maximum Score

LF maximum score was used as an index of word quality because it represents the least common word (i.e., lower frequency words possess higher frequency values, as they are rankings of frequency within the COCA) that each athlete was able to produce across all three trials of the D-KEFS VF LF subtest. This measure was significantly but weakly correlated with birth order and athlete raw and scaled score on the VF LF subtest itself (see Table 45). Regarding birth order, athletes who were earlier in the birth order, tended to produce a higher LF maximum score. This means that of the words athletes produced, earlier born children tended to achieve higher maximum LF scores, which represent less common words across the three trials of this task. Research has shown that birth order may in fact influence vocabulary size; Alfred Adler (1931, 1937) proposed that first born children tend to have larger vocabularies because they receive more individual attention from parents and have more opportunities for verbal interaction before younger siblings are born. Jenkins et al. (2015) also suggest that for similar reasons outlined by Adler (1931, 1937), first-born children may develop stronger verbal fluency skills. However, there are many factors to consider including parental interaction, family environment, and individual differences. See Table 27 for a summary of the significant relationships found for LF maximum score.

Table 27.

Letter Fluency Maximum Score Summary of Significant Findings

Letter Fluency Maximum Score	
Significant Correlations	Significant ANOVAs
Birth Order (*r* = -.18) Verbal Fluency Letter Fluency Raw Score (*r* = .17) Verbal Fluency Letter Fluency Scaled Score (*r* = .17)	None.

LF Minimum Score

LF minimum score was used as an index of word quality because it represents the most common word that athletes produced (i.e., the word with the lowest word frequency ranking in the COCA). This measure was weakly and negatively correlated with athletes' score on the visual memory composite of the ImPACT, but no other relationships were identified. This means that athletes who performed better on the ImPACT visual memory composite produced lower minimum LF scores. Ultimately, this measure is less telling than the others, as simple words are most likely to be produced on this task given that it is timed and the score is dependent on how many words are produced in one minute. See Table 28 for a summary of the significant relationship found for LF minimum score.

Table 28.

Letter Fluency Minimum Score Summary of Significant Findings

Letter Fluency Minimum Score	
Significant Correlations	Significant ANOVAs
ImPACT Visual Memory Composite (*r* = -.17)	None.

LF Average Score

LF average score was used as an index of word quality because it represents the average word frequency rankings for each athlete across all three trials of the D-KEFS VF LF subtest (i.e., of all the words each athlete produced, this is the average word frequency ranking derived from the COCA). This measure was negatively and weakly correlated with GPA, such that athletes who had a higher GPA produced lower average word frequency scores. This is interesting given the relationship between crystallized measures of intelligence and GPA, such that individuals with higher levels of crystallized intelligence tend to achieve higher GPAs in academic settings (Ackerman & Heggestad, 1997). However, given that the nature and purpose of the D-KEFS VF LF subtest is to produce as many words starting with the target letter as possible within one minute, these athletes may be using a strategy whereby they produce less sophisticated (i.e., more common) words to obtain a higher raw score in the time allotted. Otherwise, a higher LF average score was related to better performance on this task overall (VF LF raw score and scaled score) as well as on the category fluency subtest when compared to the normative sample (VF CF scaled score).

Racial identification was also related to LF average score, such that Black athletes were found to produce a higher LF average score than White athletes, meaning that the words that they produced tended to be less common and thus attain a higher word frequency score overall in the COCA. This difference could be explained by a multitude of factors, in-keeping with the AACN (2020)'s point that race is often a proxy measure for various social, economic, and environmental factors that can influence individuals' experiences and outcomes. Additionally, because it is the Corpus of American English, and language use is culturally determined, the predominantly White American culture could result in a higher representation of words produced by White individuals rather than words more commonly produced in Black American culture. Notably, Study 1 did not find a relationship between race and overall VF LF raw or scaled score, which suggests that this difference lays specifically within the quality of words produced rather than the number. See Table 29 for a summary of significant relationship found for LF average score.

Table 29.

Letter Fluency Average Score Summary of Significant Findings

Letter Fluency Average Score	
Significant Correlations	Significant ANOVAs
GPA (*r* = -.28)	Race (Black > White athletes)
Verbal Fluency Letter Fluency Raw Score (*r* = .19)	
Verbal Fluency Letter Fluency Scaled Score (*r* = .18)	
Verbal Fluency Category Fluency Scaled Score (*r* = .19)	

Overall, this study sought to identify whether the quality of the words produced on the D-KEFS verbal fluency letter fluency subtest could add to the current dissertation's exploration of factors that are important to consider in the estimation of premorbid functioning. Unlike Abeare and Seguin (2014), none of the word quality indexes developed were related to athlete performance on the TOPF, a measure of accurate oral word reading. Notably, Abeare and Seguin (2014) used the NAART, which was not used for this study because it was developed in 1989 and is thus subject to the Flynn effect, and the TOPF is a more updated word reading task. As well, Abeare and Seguin (2014) used a database of Age of Acquisition estimates, whereas this study used word frequency estimates. Although these measures have been shown to be related, they are not exactly the same and thus this leaves room for differences in findings as well. While some studies have shown that verbal fluency is related to lexical access ability and oral word reading (e.g., Levelt et al. 1999), others report that measures of phonemic fluency such as letter fluency are only weakly related to word reading (Davis et al. 2016). As such, Davis et al. (2016) suggest that caution should be exercised when extrapolating an estimate of premorbid verbal fluency abilities from measures of word reading, as well. The current study has shown that there is no significant relationship between athlete performance on verbal fluency tasks and the TOPF (Study 1) nor is there a relationship between the quality of the words produce on the D-KEFS VF LF subtest and the TOPF (Study 2).

Similar to Study 1, sample size within groups, especially minority groups, often contained few athletes and thus muddies the waters when attempting to make conclusions about these findings in a confident manner. Future research should seek to increase the number of athletes within groups and replicate these findings as such. As mentioned, the statistical power of an ANOVA is affected when there are drastically uneven sample sizes, as it results in an unfair representation of the smaller group sizes. Another limitation that is specific to verbal fluency tasks in general, is that they are conducted in English and there were at least 13 athletes whose first language was not English or English and another language. A strong predictor of verbal fluency in English is proficiency in English, especially lexical knowledge, and so athletes with greater lexical knowledge in English should generate more correct responses in the verbal fluency tasks (Paap et al., 2019). There also tend to be bilingual disadvantages in verbal fluency, such that there is cross-language interference between the two languages (Sandoval et al., 2010). Despite all of this, Study 1 found significant differences between English speakers and English+Other language speakers and French speakers, but not for the letter fluency subtest, and language was not found to have a significant relationship with word quality produced. So, despite research suggesting that there could be an influence, the disadvantage was not seen on the letter fluency subtest in terms of total performance and word quality produced.

In conclusion, the current study investigated the relationship between word quality of the words produced on the D-KEFS VF LF subtest and the TOPF, to determine whether word quality is an appropriate measure of premorbid functioning. No significant relationship was found, nor was it found in Study 1 between raw and scaled scores on the D-KEFS VF LF subtest and the TOPF. Few significant relationships were found among personal and demographic variables, and those that were identified were weak in strength of correlation or possessed an unequal sample size that affected the power of the comparison. Although this study demonstrates that there are some significant but weak relationships with performance on the D-KEFS VF LF subtest and the quality of the words produced, it promotes the idea that there are many reasons for a person's performance on this task, and like Study 1, highlights the use of a strong biopsychosocial approach to neuropsychological assessment.

CHAPTER 5

Study III: A Regression-Based Approach to Estimating Premorbid Functioning

In the past, strategies have attempted to use both demographic variables and test data to predict premorbid functioning via regression formulae. Some of the first to do this did so by developing a regression formula based on the variables of age, gender, race, occupation, education, urban vs. rural settings, and region (Barona et al., 1984). Other examples of this strategy are the Oklahoma Premorbid Intelligence Estimate – 3 (OPIE-3; Schoenberg et al., 2006), Hopkins Adult Reading Test (HART; Schretlen et al., 2009), and Wechsler Test of Adult Reading (WTAR; Holdnack, 2001). These methods use different combinations of demographic variables as well as objective test scores to predict IQ scores. These methods are useful because they can provide unbiased estimates of premorbid functioning and can be modified in light of additional qualitative information available to the clinician (Crawford & Allan, 1997). However, it is important to remember that regression to the mean affects all these methods, and often underestimates individuals' IQs who are above 125 and over-estimate individual IQs that are below 75 (Barona et al., 1984; Reynolds, 1997).

Study Aims and Hypothesis

The current study aimed to use the results of Study 1 and 2 to generate prediction algorithms using standard multiple regression in which subtest raw scores and demographic variables with the strongest relationships to each test are predictors of neuropsychological test scores. Unfortunately, Study 1 and 2 did not provide strong evidence for the use of these variables in regression formulae, as many correlations were deemed to be weak. Rather than estimating IQ, like much of the past research has, this study aims to attempt to estimate more specific test scores, given that one's intelligence is composed of numerous different skills and abilities. Additionally, in Study 1 and 2, the findings show different variables having different relationships to each test score. As such, Study 3 aimed to attempt to create regression-based methods of estimating premorbid functioning that is informed by the findings from Study 1 and 2.

The optimization and improvement of the estimation of premorbid functioning in the sport neuropsychology world is important and necessary. Baseline testing is often used to help operationalize the meaning of "fully recovered" for athletes after sustaining a concussion (Piland et al., 2010), but the utility of the comparison between post-injury and baseline test data in return-to-play decisions is based heavily on the integrity of the baseline data (Erdal, 2012). Prevalence of invalid performance during baseline testing has been estimated to be more common than initially thought (Abeare et al., 2018; Messa et al., 2020), and there is insufficient evidence that having baseline test results to compare to is superior to not having baseline test results in the sporting world (Iverson & Schatz, 2014). Therefore, the use of regression-based methods using multiple relevant predictor variables could improve the accuracy of premorbid functioning estimates beyond current approaches. Given that relationships were found to be mostly weak in Study 1 and 2, it is unlikely that these approaches will be optimal for use in this population, which emphasizes that better understanding these variables' contribution to the model is useful in moving towards a more informed neuropsychological assessment generally. If successful, this approach could act as a motivator for to the implementation of shorter baseline assessments or could act as the first step towards eliminating baseline testing entirely.

Method

Procedure

Following the general procedure described previously and the completion of Study 1 and Study 2, significant relationships (i.e., correlation and ANOVA findings) for the ImPACT composite scores, the Trail Making Test A and B raw scores, and the D-KEFS verbal fluency subtest raw scores were evaluated. The TOPF was not included, as it is already considered to be a measure of premorbid functioning.

Data Analysis

All analyses were conducted using IBM SPSS Statistics v.29.0.0.0. Prior to conducting the primary analyses, the assumptions of standard multiple regression were considered. These include assumptions of linearity (the relationship between the predictors and dependent variables are linear), independence of observations (the values of one observation should not be influenced by or dependent on the values of other observations), homoscedasticity (the variability of the residuals should be constant across all levels of the independent variables), normality of the residuals (the residuals or errors should follow a normal distribution), no perfect multicollinearity (no perfect linear relationships among the independent variables), no autocorrelation (the residuals should not be correlated with each other), and additivity (changes in one predictor variable do not depend on the values of other predictor variables).

A power analysis was performed using G*Power version 3.1.9.6 (Faul et al., 2007) to determine the maximum number of predictor variables that could be included based on the sample size obtained for this study (n = 158). Results indicated that for the required sample size to achieve 80% power for detecting a medium effect (Cohen, 1988), at a significance criterion of α = .05, up to 20 predictor variables could be included. For the required sample size to achieve 95% power for detecting a medium effect, a maximum of 7 predictor variables could be included. While adding more predictors to a regression model can improve its explanatory power and capture complex relationships, it is essential to strike a balance and consider the potential drawbacks such as overfitting and multicollinearity. Additionally, many relationships outlined in Study 1 were found to be weak in nature. Therefore, variable selection was completed with these considerations, on a test-by-test basis.

Results

The Immediate Post-Concussion Assessment and Cognitive Testing (ImPACT)

Verbal Memory Composite. The significant relationships identified for the verbal memory composite score in Study 1 were considered (grade 7/8 science grades, ImPACT visual memory composite, ImPACT visual motor speed composite, and ImPACT impulse control composite). ImPACT composite scores that were identified as having a relationship with the ImPACT verbal memory composite were not included, as the ImPACT is most often administered in its entirety and therefore obtaining the other composite scores would include the ImPACT verbal memory composite as well. Therefore, the only significant relationship for inclusion is athletes' self-reported grade 7/8 science grades.

Because there was only one predictor variable identified for the ImPACT verbal memory composite, the predictive relationship between ImPACT verbal memory composite and grade 7/8 self-reported science grade was examined using a simple linear regression model. Notably, there is no evidence to suggest that one's grade 7/8 self-reported science grade is related to verbal memory ability in the literature, and the correlation was deemed to be weak in Study 1. Additionally, both the verbal memory composite and grade 7/8 science grades were not normally distributed. The linear regression analysis revealed a statistically significant relationship $R^2 = .07$, $F (1, 147) = 10.54$, $p = .001$. Therefore, only 7% of the variance of the ImPACT verbal memory composite is explained by the variance of grade 7/8 self-reported science grades (Table 30).

Table 30.

ImPACT Verbal Memory Regression Coefficient Table

	Unstandardized B	Coefficients Standard Error	Standardized Coefficients Beta	t	Significance
Constant	90.54	1.29		70.30	< .001
Grade 7/8 Science	-1.06	.33	-.26	-3.25	.001

Visual Memory Composite. The significant relationships identified for the visual memory composite score in Study 1 were considered (grade 7/8 math grades, grade 7/8 science grades, ImPACT verbal memory composite score, ImPACT visual motor speed composite score, ImPACT reaction time composite score, ImPACT impulse control composite score, Trails B raw score, Trails B T score). ImPACT composite scores that were identified as having a relationship with the ImPACT visual memory composite were not included. Therefore, the significant relationships for inclusion are athletes' self-reported grade 7/8 science and math grades, Trails B raw score, Trails B T score, and possession of a scholarship. Notably, grade 7/8 math and science, and Trails B raw and T score exhibited multicollinearity, and so only grade 7/8 math grades and Trails B raw score will be included. This is because grade 7/8 math grades had a higher correlation with the ImPACT Visual Memory Composite and the Trails B raw score is a score with more utility for these purposes since it is not already normed.

The predictive relationship between ImPACT visual memory composite and these predictors was examined using a multiple linear regression model. Notably, all correlations were deemed to be weak in Study 1. Additionally, all variables were not normally distributed, and multicollinearity was not present ($r < .70$). The linear regression analysis revealed a statistically significant relationship $R^2 = .18$, $F(3, 146) = 10.81$, $p = < .001$. Therefore, 18% of the variance of the ImPACT visual memory composite is explained by the included predictor variables. The individual predictors were examined further and indicated that grade 7/8 math ($t = -2.08$, $p = .04$), Trail B raw score ($t = -4.05$, $p < .001$), and scholarship attainment ($t = 2.28$ $p = .02$) were all significant predictors. See Table 31 for a summary of the predictor coefficients.

Table 31.

ImPACT Visual Memory Regression Coefficient Table

	Unstandardized B	Coefficients Standard Error	Standardized Coefficients Beta	t	Significance
Constant	84.42	2.29		36.86	< .001
Grade 7/8Math	-.767	.368	-.160	-2.08	.039
Trails B Raw Score	-.131	.032	-.311	-4.05	< .001
Award	5.92	2.60	.171	2.28	.024

Visual Motor Speed Composite. The significant relationships identified for the visual motor speed composite score in Study 1 were considered (GPA, grade 7/8 math grades, grade 7/8 science grades, grade 7/8 language grades, grade 7/8 social studies grades, ImPACT verbal memory composite, ImPACT reaction time composite, ImPACT visual memory composite, Trails A raw and T score, Trails B raw and T score, VF-LF raw and scaled score, VF-CF raw and scaled score, VF-CS # raw and scaled score, VF-CS accuracy raw and scaled score, sex, and sport). ImPACT composite scores that were identified as having a relationship with the ImPACT visual motor speed composite were not included. Therefore, the significant relationships for inclusion are GPA, grade 7/8 math, science, social studies, and language grades, Trails A raw and T score, Trails B raw and T score, VF-LF raw and scaled score, VF-CS # correct raw and scaled score, VF-CS accuracy raw and scaled score, sex, and sport. Notably, grade 7/8 math, language, science, and social studies scores exhibited multicollinearity, and so only grade 7/8 math grades were included due to the highest correlation with the ImPACT visual motor speed composite. Additionally, only Trails A and B raw scores were included, due to multicollinearity with their respective T scores. The same rule applied for the verbal fluency subtests, whereby only raw scores were included, and VF-CS accuracy raw score was removed due to multicollinearity with VF-CS number correct.

For the purposes of the regression equation, sport was dummy coded to be included. The linear regression analysis revealed a statistically significant relationship $R^2 = .38$, F (9, 97) = 6.65, $p < .001$. Therefore, 38% of the variance of the ImPACT visual motor speed composite is explained by the included predictor variables. The individual predictors were examined further and indicated that sport ($t = -.52$, $p = .61$), GPA ($t = .14$, $p = .89$) grade 7/8 math ($t = -.87$, $p = .39$), Trails A raw score ($t = -.90$, $p = .37$), VF-LF raw score ($t = .79$, $p = .43$), VF-CS number correct ($t = .35$, $p = .73$), and sex ($t = .46$, $p = .46$) were not significant predictors of the model. Trails B raw score ($t = -4.99$, $p < .001$)

and VF-CF raw score (t = 2.16, p = .03) were significant predictors. As such, Trails B raw score and VF-CF raw score accounted for 31% of the variance of the ImPACT visual motor speed composite, R^2 = .31, F (2, 153) = 33.72, p < .001. See Table 32 for a summary of the predictor coefficients.

Table 32.

ImPACT Visual Motor Speed Composite Regression Coefficients

	Unstandardized B	Coefficients Standard Error	Standardized Coefficients Beta	t	Significance
Constant	43.02	10.82		3.98	< .001
GPA	.01	.08	.01	.14	.89
Grade 7/8Math	-.19	.22	-.08	-.87	.39
Trails A Raw Score	-.06	.07	-.08	-.89	.37
Trails B Raw Score	-.11	.02	-.45	-4.99	< .001
VF-LF Raw Score	.05	.06	.07	.79	.43
VF-CF Raw Score	.15	.07	.22	2.16	.03
VF-CS #Correct Raw Score	.09	.27	.03	.35	.73
Sex	-2.25	3.05	-.16	-.74	.46
Sport	-.27	.53	-.11	-.52	.61

Reaction Time Composite. The significant relationships identified for the reaction time composite score in Study 1 were considered (GPA, grade 7/8 math grades, grade 7/8 language grades, grade 7/8 science grades, ImPACT visual memory composite, and sport) and are outlined in Table 31. ImPACT composite scores that were identified as having a relationship with the ImPACT reaction time composite were not included. Therefore, the significant relationships for inclusion are GPA, grade 7/8 math, science, and language grades, Trails A raw and T score, Trails B raw and T score, VF-LF raw and scaled score, VF-CF raw and scaled score, VF-CS # correct raw and scaled score, VF-CS accuracy raw and scaled score, sex, and sport. As was done for the visual motor speed composite, only grade 7/8 math grades and raw scores of neuropsychological tests were included, minus the VF-CS accuracy raw score.

For the purposes of the regression equation, sport was dummy coded to be included. The linear regression analysis revealed a statistically significant relationship $R^2 = .24$, $F(8, 98) = 3.85$, $p < .001$. Therefore, 24% of the variance of the ImPACT reaction time composite is explained by the included predictor variables. The individual predictors were examined further and indicated that GPA ($t = -.31$, $p = .76$), grade 7/8 math ($t = 1.01$, $p = .32$), Trails B raw ($t = 1.54$, $p = .13$), VF-LF raw ($t = -.43$, $p = .67$), VF-CS number correct raw ($t = .71$, $p = .48$), sport ($t = -1.59$, $p = .11$) and sex ($t = .45$, $p = .66$) were not significant predictors of the model. Trails A raw score ($t = 2.70$, $p = .01$) and VF-CF raw score ($t = -2.25$, $p = .03$) were significant predictors. As such, Trails A raw score and VF-CF raw score accounted for 16% of the variance of the ImPACT reaction time composite, $R^2 = .16$, $F(2, 154) = 14.20$, $p < .001$. See Table 33 for a summary of the predictor coefficients.

Table 33.

ImPACT Reaction Time Composite Regression Coefficients

	Unstandardized B	Coefficients Standard Error	Standardized Coefficients Beta	t	Significance
Constant	.59	.15		4.07	< .001
GPA	.00	.00	-.03	-.31	.76
Grade 7/8 Math	.00	.00	.10	1.01	.32
Trails A Raw Score	.00	.00	.26	2.70	.01

Trails B Raw Score	.00	.00	.15	1.54	.13
VF-LF Raw Score	.00	.00	-.04	-.43	.67
VF-CF Raw Score	.00	.00	-.25	-2.25	.03
VF-CS # Correct Raw Score	.00	.00	.07	.71	.48
Sex	.01	.02	.05	.45	.48
Sport	-.01	.01	-.38	-1.59	.11

Impulse Control Composite. The significant relationships identified for the impulse control composite score in Study 1 were considered (ImPACT verbal memory composite, ImPACT visual memory composite). Apart from other ImPACT composite scores, no other relationships were found, and so this composite score could not be included in this portion of the analyses.

The Trail Making Test (TMT)

Trail Making Test A (TMT-A). The significant relationships identified for the Trail Making Test A raw scores were considered (ImPACT visual motor speed composite score, ImPACT reaction time composite score, Trails B raw and T score, VF-CF raw and scaled score, and sex). Significant relationships for consideration include the ImPACT visual motor speed and reaction time composite scores, Trails B raw and T scores, VF-CF raw and scaled scores, and sex. Because the TMT-A T score relationships only included one difference, years of education instead of sex, only the raw score was evaluated because the T score has been normed based on age, race, sex, and years of education. Additionally, as mentioned previously, only VF-CF raw score is included.

The linear regression analysis revealed a statistically significant relationship R^2 = .19, F (5, 150) = 6.95, *p* < .001. Therefore, 19% of the variance of the Trail Making Test A raw score is explained by the included predictor variables. The individual predictors were examined further and indicated that the ImPACT visual motor speed composite (*t* = -1.07, *p* = .29), VF-CF raw score (*t* = -.67, *p* = .50), and sex (*t* = .72, *p* = .47) were not significant predictors of the model, but ImPACT reaction time composite (*t* = 2.41, *p* = .02) and Trails B raw score (*t* = 2.57, *p* = .01) were significant predictors. As such, Trails B raw score and ImPACT reaction time composite score accounted for 17% of the variance of Trails A raw score, R^2 = .17, F (2, 154) = 15.78, *p* < .001. See Table 34 for a summary of the predictor coefficients.

Table 34.

Trail Making Test A Raw Score Regression Coefficients

	Unstandardized B	Coefficients Standard Error	Standardized Coefficients Beta	t	Significance
Constant	15.38	7.92		1.94	.05
Visual Motor Speed Composite	-.12	.11	-.10	-1.07	.29
Reaction Time Composite	16.32	6.78	.20	2.41	.02
Trails B Raw Score	.06	.02	.22	2.57	.01
VF-CF Raw Score	-.04	.06	-.05	-.67	.50
Sex	.87	1.21	.06	.72	.47

Trail Making Test B (TMT-B). The significant relationships identified for the Trail Making Test B raw scores were considered (GPA, grade 7/8 math grades, grade 7/8 language grades, grade 7/8 science grades, grade 7/8 social studies grades, father/second parent’s education, SES ladder, ImPACT visual memory composite score, ImPACT visual motor speed composite score, ImPACT reaction time composite score, Trails A raw and T score, VF-LF raw and scaled score, VF-CF raw and scaled score, VF-CS accuracy raw and scaled score). Significant relationships for consideration include GPA, grade 7/8 math, language, science, and social studies grades, father/second parent's education, SES impression, ImPACT visual memory composite, visual motor speed composite, and reaction time composite, Trails A raw and T score, VF-LF raw and scaled score, VF-CF raw and scaled score, and VF-CS number correct raw and scaled score. As mentioned previously, only grade 7/8 math grades (and not language, science, or social studies), and neuropsychological test raw scores are included.

The linear regression analysis revealed a statistically significant relationship R^2 = .49, F (11, 82) = 7.36, p < .001. Therefore, 49% of the variance of the Trail Making Test B raw score is explained by the included predictor variables. The individual predictors were examined further and indicated that GPA (t = -.32, p = .75), grade 7/8 math (t = -.52, p = .60), SES impression (t = -.54, p = .59), ImPACT reaction time composite (t = -.49, p = .63), Trails A raw score (t = .96, p = .34), VF-LF raw score (t = 1.86, p = .07), VF-CF raw score (t = 1.23, p = .22), and VF-CS number correct raw score (t = -.55, p = .58) were not significant predictors of the model, but Parent 2/father's education (t = -.237, p = .02), ImPACT visual memory composite (t = -4.83, p < .001), and ImPACT visual motor speed composite (t = -4.83, p < .001) were significant predictors. As such, these three variables accounted for 37% of the variance of Trails B raw score, R^2 = .37, F (3, 138) = 26.45, p < .001. See Table 35 for a summary of the predictor coefficients.

Table 35.

Trail Making Test B Raw Score Regression Coefficients

	Unstandardized B	Coefficients Standard Error	Standardized Coefficients Beta	t	Significance
Constant	226.20	44.41		5.09	< .001
GPA	-.11	.34	-.03	-.32	.75
Grade 7/8 Math	-.52	.99	-.05	-.52	.60
Parent 2/Father Education	-2.79	1.18	-.20	-2.37	.02
Impression of SES	-.75	1.38	-.04	-.54	.59
Visual Memory Composite	-.72	.19	-.34	-3.72	< .001
Visual Motor Speed Composite	-2.08	.43	-.49	-4.83	< .001
Reaction Time Composite	-12.53	25.85	-.05	-.49	.63
Trails A Raw Score	.33	.34	.09	.96	.34
VF-LF Raw Score	.45	.24	.17	1.86	.07
VF-CF Raw Score	.38	.31	.13	1.23	.22
VF-CS Number Correct Raw Score	-.60	1.09	-.05	-.55	.58

D-KEFS Verbal Fluency

D-KEFS Verbal Fluency – Letter Fluency. The significant relationships identified for the D-KEFS verbal fluency letter fluency raw scores were considered (years of education, TOPF raw score, ImPACT visual motor speed composite, ImPACT reaction time composite, Trails B raw score, VF-CF raw and scaled score, VF-CS number correct raw and scaled score, CF-CS accuracy raw and scaled score). Significant relationships for consideration include years of education, TOPF raw score, ImPACT visual motor speed and reaction time composite, trails B raw score, VF-CF raw and scaled score, VF-CS number correct raw and scaled score, and VF-CS accuracy raw and scaled score. As mentioned previously, neuropsychological test scaled scores were removed, as was VS-CS accuracy raw score.

The linear regression analysis revealed a statistically significant relationship R^2 = .31, F (7, 148) = 9.34, *p* < .001. Therefore, 31% of the variance of VF-LF raw score is explained by the included predictor variables. The individual predictors were examined further and indicated that TOPF raw score (*t* = 1.07, *p* = .29), ImPACT visual motor speed composite score (*t* = 1.39, *p* = .17), ImPACT reaction time composite score (*t* = -.40, *p* = .69), Trails B raw score (*t* = .49, *p* = .63), and VF-CS number correct raw score (*t* = .03, *p* = .98) were not significant predictors of the model, but years of education (*t* = 2.60, *p* = .01) and VF-CF raw score (*t* = 5.28, *p* < .001) were significant predictors. As such, these two variables accounted for 28% of the variance for the VF-LF raw score, R^2 = .28, F (2, 154) = 30.59, *p* < .001. See Table 36 for a summary of the predictor coefficients.

Table 36.

Verbal Fluency – Letter Fluency Raw Score Regression Coefficients

	Unstandardized B	Coefficients Standard Error	Standardized Coefficients Beta	t	Significance
Constant	-8.40	11.57		-73	.47
Years of Education	1.25	.48	.19	2.60	.01
TOPF Raw Score	.07	.07	.08	1.07	.29
Visual Motor Speed Composite	.20	.14	.12	1.39	.17
Reaction Time Composite	-3.33	8.38	-.03	-.40	.69
Trails B Raw Score	.01	.03	.04	.49	.63

VF-CF Raw Score	.45	.09	.44	5.28	< .001
VF-CS Raw Score	.01	.30	.002	.03	.98

D-KEFS Verbal Fluency – Category Fluency. The significant relationships identified for the D-KEFS verbal fluency category fluency raw scores were considered (ImPACT visual motor speed composite, ImPACT reaction time composite, Trails A raw and T score, Trails B raw and T score, VF-CS number correct raw and scaled score, VF-CS accuracy raw and scaled score, and first language. Significant relationships for consideration include the ImPACT visual motor speed and reaction time composite, trails A raw and T score, Trails B raw and T score, VF-CS number correct raw and scaled score, VF-CS accuracy raw and scaled score, and first language. As mentioned previously, neuropsychological test T and scaled scores were removed, as was VF-CS accuracy raw score.

The linear regression analysis revealed a statistically significant relationship $R^2 = .32$, F (6, 149) = 11.70, $p < .001$. Therefore, 32% of the variance of VF-CF raw score is explained by the included predictor variables. The individual predictors were examined further and indicated that ImPACT reaction time composite score ($t = -1.55$, $p = .13$), trails A raw score ($t = -.51$, $p = .61$), and first language ($t = -1.57$, $p = .12$) were not significant predictors of the model, but ImPACT visual motor speed composite ($t = 2.44$, $p = .02$), trails B raw score ($t = 2.21$, $p = .03$), and VF-CS number correct raw score ($t = 5.83$, $p < .001$) were significant predictors. As such, these variables accounted for 29% of the variance for the VF-CF raw score, $R^2 = .29$, F (3, 152) = 21.07, $p < .001$. See Table 37 for a summary of the predictor coefficients.

Table 37.

Verbal Fluency – Category Fluency Raw Score Regression Coefficients

	Unstandardized B	Coefficients Standard Error	Standardized Coefficients Beta	t	Significance
Constant	18.64	9.49		1.96	.05
Visual Motor Speed Composite	.32	.13	.21	2.44	.02
Reaction Time Composite	-12.58	8.15	-.12	-1.55	.13
Trails A raw score	-.05	.10	-.04	-.51	.61
Trails B raw score	.06	.03	.17	2.21	.03
VF-CS Number Correct raw score	1.49	.26	.42	5.83	< .001
First Language	-.82	.52	-.11	-1.57	.12

D-KEFS Verbal Fluency – Category Switching Number Correct. The significant relationships identified for the D-KEFS verbal fluency category switching number correct raw scores were considered (GPA, grade 7/8 social studies grades, TOPF raw score, ImPACT visual motor speed composite, ImPACT reaction time composite, Trails B raw score, VF-LF raw and scaled score, VF-CF raw and scaled score, VF-CS accuracy raw and scaled score, sex, race, and first language). Significant relationships for consideration include GPA, grade 7/8 social studies grades, TOPF raw score, ImPACT visual motor speed and reaction time composites, Trails B raw score, VF-LF raw and scaled score, VF-CF raw and scaled score, race, sex, and first language. Notably, VF-CS accuracy raw and scaled score are related, however these scores are from the same task and so completing D-KEFS VF CS results in both the number correct and accuracy scores. As was done previously, only neuropsychological test raw scores are included.

The linear regression analysis revealed a statistically significant relationship $R^2 = .30$, F (11, 94) = 3.73, $p < .001$. Therefore, 30% of the variance of VF-CS number correct raw score is explained by the included predictor variables. The individual predictors were examined further and indicated that only VF-CF raw score contributed significantly to the model (t = 4.07, p < .001). As such, VF-CF raw score accounted for 24% of the variance for the VF-CS number correct raw score, $R^2 = .24$, F (1, 155) = 48.93, p < .001. See Table 38 for a summary of the predictor coefficients.

Table 38.

Verbal Fluency – Category Switching Number Correct Raw Score Regression Coefficients

	Unstandardized B	Coefficients Standard Error	Standardized Coefficients Beta	t	Significance
Constant	5.24	4.20		1.25	.22
GPA	.04	.03	.12	1.14	.26
Grade 7/8 Social Studies	.04	.12	.03	.31	.76
TOPF raw	-.004	.02	-.02	-.19	.85
Visual Motor Speed Composite	.03	.04	.08	.73	.47
Reaction Time Composite	1.48	2.35	.06	.63	.53
Trails B Raw	-.01	.01	-.09	-.86	.39
VF-LF Raw	.02	.02	.07	.65	.52
VF-CF Raw	.10	.03	.44	4.07	< .001
Sex	-.63	.50	-.13	-1.26	.21
Race	.11	.12	.09	.90	.37
First Language	-.02	.16	-.01	-.13	.90

D-KEFS Verbal Fluency – Category Switching Accuracy. The significant relationships identified for the D-KEFS verbal fluency category switching accuracy raw scores were considered (grade 7/8 social studies grades, ImPACT visual motor speed composite, ImPACT reaction time composite, VF-LF raw and scaled score, VF-CF raw and scaled score, VF-CS number correct raw and scaled score, sex, and other diagnosis). Significant relationships for consideration include grade 7/8 social studies grades, ImPACT visual motor speed and reaction time composite, VF-LF raw and scaled score, VF-CF raw and scaled score, sex, and other diagnosis. As mentioned previously, the VF-CS number correct score is not included, as VF-CS number correct and VF-CS accuracy are highly correlated. Additionally, neuropsychological test scaled scores are not included.

The linear regression analysis revealed a statistically significant relationship R^2 = .25, F (7, 141) = 6.59, $p < .001$. Therefore, 25% of the variance of VF-CS accuracy raw score is explained by the included predictor variables. The individual predictors were examined further and indicated that only VF-CF raw score contributed significantly to the model ($t = 3.56$, $p < .001$). As such, VF-CF raw score accounted for 19% of the variance for the VF-CS accuracy raw score, R^2 = .19, F (1, 155) = 35.41, $p < .001$. See Table 39 for a summary of the predictor coefficients.

Table 39.

Verbal Fluency – Category Switching Accuracy Raw Score Regression Coefficients

	Unstandardized B	Coefficients Standard Error	Standardized Coefficients Beta	t	Significance
Constant	10.58	2.98		3.55	< .001
Grade 7/8 Social Studies	-.13	.11	-.09	-1.13	.26
Visual Motor Speed Composite	.04	.04	.07	.82	.41
Reaction Time Composite	-3.02	2.58	-.10	-1.17	.24
VF-LF raw score	.02	.03	.06	.64	.52

VF-CF raw score	.10	.03	.32	3.56	< .001
Sex	-.85	.50	-.14	-1.68	.10

Discussion

The purpose of the Study 3 was to incorporate results from Study 1 and Study 2 to generate prediction algorithms using standard simple and multiple linear regression, in which subtest raw scores and demographic variables that were shown to have a relationship with test scores were used as predictor variables. Therefore, Study 3 attempted to optimize and improve past regression-based methods of estimating premorbid functioning, by focusing on specific neuropsychological test scores rather than IQ more generally, and by personalizing the approach to the sport neuropsychology setting.

Study 1 found mostly weak correlations between test scores and demographic, personal, and educational variables, and so developing regression equations with high predictive value was difficult. When models are being developed for prediction, a higher R-squared value is more desirable, however a low R-squared value could still be valuable if it contributes to an understanding of the relationships between variables. Therefore, for the purposes of this study, predicting relationships was not necessarily feasible but better understanding these variables for the purpose of conducting better-informed neuropsychological assessments was more appropriate. This approach is in-keeping with the AACN (2020)'s position statement, reflecting the importance of a "precision-medicine" approach, and thus this study contributes to that movement.

Regression analyses were completed for the ImPACT composite scores, the Trail Making Test A and B, and for the D-KEFS Verbal Fluency subtasks. The TOPF was not included in these analyses because the purpose of the TOPF itself is to estimate premorbid functioning, and so attempting to generate prediction models for a test whose main purpose is predicting premorbid functioning is unnecessary and redundant. As such, the ImPACT, Trail Making Test, and D-KEFS Verbal Fluency are discussed in more detail.

The ImPACT

Across ImPACT composite scores (Verbal Memory, Visual Memory, Visual Motor Speed, Reaction Time, Impulse Control), fit of the regression models were low. R^2 is typically defined as how well the regression model fits the observed data values, and how well it explains the fitted data. In other words, the closer the R^2 value is to 1, the more accurate the model. It was determined that the Visual Motor Speed composite had the best fit using Trails B raw score and VF-CF raw score as predictor variables, with the strength of the relationship being 31%. As mentioned, a higher R^2 value is desirable for predictive models, and thus the takeaway from these findings is that Visual Motor Speed could be explained by one's scores on Trails B and VF-CF to some extent, but not enough to develop a predictive equation for these purposes. See Table 40 for a summary of these findings.

Table 40.

ImPACT Regression Summary

ImPACT Composite	Contributing Predictors	R^2
Verbal Memory	Grade 7/8 science	0.07 (7%)
Visual Memory	Trails B Raw Score Scholarship Attainment	0.18 (18%)
Visual Motor Speed	Trails B Raw Score VF-CF Raw Score	0.31 (31%)
Reaction Time	Trails A Raw Score VF-CF Raw Score	0.16 (16%)

Impulse Control	N/A	N/A

Conceptually, it makes sense that individuals who perform better on Trails B also perform better on the ImPACT visual motor speed composite, as both tasks involve visual processing, processing speed, and motor coordination to some extent. VF-CF's relationship to these tasks could also be at least partially explained by the fact that semantic fluency tasks engage a distributed network of brain regions involved in language processing, executive functions, and cognitive control. Semantic fluency is also involved in lexical access speed (Shao et al., 2014), which can help explain its relationship to Trails B (accessing letters in alphabetical order) and the Visual Motor Speed composite, which is composed of two tasks which require access of letters (X's and O's), and matching numbers to symbols (Symbol Match).

The Trail Making Test

The Trail Making Test (TMT) A and B also exhibited poor fit for predictive purposes, with TMT-B having better fit than TMT-A (see Table 41). For TMT-B, significant predictor variables included father's education, ImPACT Visual Memory Composite, and ImPACT Visual Motor Speed Composite, explaining 37% of the variance for TMT-B.

Table 41.

Trail Making Test Regression Summary

TMT Condition	Contributing Predictors	R^2
Trail Making Test A	Trail Making Test B Raw Score ImPACT Reaction Time Composite	0.17 (17%)

Trail Making Test B	Parent 2/Father's Education ImPACT Visual Memory Composite ImPACT Visual Motor Speed Composite	0.37 (37%)

Ultimately, the contribution of the ImPACT visual memory and visual motor speed composites make sense, as the TMT-B is a task of speeded visual-motor ability, as well as executive functioning skills. The contribution of the father's education is also understandable, given that higher parental education is associated with better school achievement and higher intelligence of the child (Tamayo Martinez et al., 2022). In recent generations, however, mothers' and fathers' education has been shown to equally influence children's intellectual abilities (Cave et al., 2022).

D-KEFS Verbal Fluency

The verbal fluency subtests also exhibited poor predictive power, and significant predictors often included other neuropsychological test scores. As mentioned previously, a relationship between VF-CF and the visual motor speed composite and TMT-B raw score is evident. VF-CS number correct raw score was also a significant predictor, which makes sense given that it is a similar task relying on categorical knowledge. As such, these predictors explained 29% of the variance for VF-CF. 28% of the variance for VF-LF was explained by years of education and VF-CF raw score. Years of education make sense as well, as VF-LF is a word generation task and typically those who have completed more education are exposed to more words. Education has been shown to be the best predictor of performance on both phonological and semantic verbal fluency tasks

(Lubrini et al., 2022). See Table 42 for a summary of the regression findings for the D-KEFS verbal fluency subtests.

Table 42.

Verbal Fluency Regression Summary

Verbal Fluency Condition	Contributing Predictors	R^2
Verbal Fluency – Letter Fluency	Years of education VF-CF raw score	0.28 (28%)
Verbal Fluency – Category Fluency	Visual Motor Speed Composite Trails B raw score VF-CS number correct raw score	0.29 (29%)
Verbal Fluency – Category Switching Number Correct	VF-CF raw score	0.24 (24%)
Verbal Fluency – Category Switching Accuracy	VF-CF raw score	0.19 (19%)

Overall, the most salient limitation in our ability to predict premorbid functioning in specific cognitive domains was that the regression equations were created using variables that were shown to have only weak correlations in Study 1, and Study 2 word frequency values were not significantly related and thus were not included. Therefore, developing regression equations with high predictive value was difficult and limited. Nonetheless, this study aimed to provide sport neuropsychologists with a richer understanding of the interrelationships of these variables for the purpose of conducting better-informed neuropsychological assessments. As well, regression analyses are often based on the assumption that the relationships that are observed in the data can generalize to the population from which the data was drawn. It is important to note that this sample, while of a sufficient total sample size, was limited in terms of its representation of different Racial groups, languages, and birth countries. In addition, one must consider that there could be other factors influencing the relationships between independent and dependent variables that were not accounted for during this assessment (e.g., negative mental health symptoms, sleep, pain, etc).

In conclusion, while this study is not able to provide strong evidence of the predictive power of the included demographic, personal, and educational variables in determining test performance, it does provide sport neuropsychologists with a better understanding of the possible factors to consider when making return-to-play decisions for this unique population (contributing predictor variables included: scholarship attainment, grade 7/8 science self-reported grades, father's education, years of education).

CHAPTER 6

General Discussion

The current dissertation aimed to improve our understanding of the estimation of premorbid intellectual functioning in sport neuropsychology. Currently, neuropsychological assessment of sport-related concussion is lacking in this knowledge base and in approaches that emphasize the importance of considering personal, educational, and demographic factors when working with this unique population. Accurate estimation of premorbid functioning in athletics is a complicated issue but is especially crucial because incorrect estimation could result in negative consequences, including an athlete prematurely returning to play and experiencing negative long-term outcomes associated with multiple concussion sustained in close time proximity. Additionally, baseline testing is a resource-intensive and time-consuming process with many identified shortcomings. The results of this study provide a complex picture regarding best estimates of premorbid functioning in sport-related concussion. The identified predictors of premorbid functioning can inform neuropsychologists' approach to baseline testing.

The first study acted as a general exploration of the measures and variables that best relate to measures deemed to be useful in the estimation of premorbid functioning, specifically in a varsity athletics setting. This was the first known study to incorporate such a wide range of demographic, personal, and educational variables in estimating premorbid functioning. It was hypothesized that the strongest relationships with baseline cognitive functioning would be found between tests of crystallized intelligence and educational demographic variables, including the TOPF, years of education, quality of education, and parental education.

Across the cognitive tests included in the test battery, several significant correlations were found and ranged from weak to strong. Predictably, strong correlations were found between test scores measuring related cognitive abilities within the cognitive test battery, whereas correlations between the demographic, personal, and education variables with test scores tended to be weak. Variables with significant correlations included: parental education, full vs. part-time studies, possession of a scholarship, grade 7/8 self-reported grades, current GPA, sex, years of education, handedness, and impression of SES. Other significant relationships that were identified lacked an appropriate sample size, and thus should be further explored (i.e., race, sport, and first language). Overall, this study gives neuropsychologists an idea of which variables that could be related to test performance. It promotes the assessment of a wider range of variables than are traditionally considered. It also highlights the idea that no one or two influential variables should be used to estimate premorbid functioning for the purposes of making important return-to-play decisions.

The second study aimed to replicate and expand upon Abeare and Seguin's (2014) findings that people with a higher NAART estimated IQ are more likely to generate words on a phonemic fluency task with a higher Age of Acquisition (AoA). As such, this study worked to determine whether the quality of athlete responses on the D-KEFS Verbal Fluency Test – Letter Fluency subtest (VF-LF) was related to estimates of premorbid functioning provided by the TOPF. It was hypothesized that much like in Abeare and Seguin's (2014) study, word qualities related to AoA would be a strong predictor of premorbid functioning and would then account for unique variance in the prediction of premorbid functioning in Study 3.

Through the use of word frequency values derived from the Corpus of Contemporary American English, the letter fluency maximum score (i.e., the lowest frequency and therefore least common word that each athlete produced) was significantly but weakly correlated with birth order, whereby athletes who were earlier in the birth order tended to produce a higher score (i.e., less common word), which is in-keeping with Adler (Jenkins et al., 2015) who suggested that first-born children may develop stronger verbal fluency skills. When looking at athlete average LF word frequency score (i.e., of all of the words that each athlete produced, this is the average word frequency ranking derived from the COCA), this measure was negatively and weakly correlated with GPA. This relationship showed that those who had a higher GPA produced more common word scores and may be related to a strategy to produce more common and frequently encountered words in order to excel at the task itself, which was to produce as many words as quickly as possible that started with a specific letter.

Racial identification was also related to LF average frequency score, where Black athletes were found to produce a less common words than White athletes (i.e., words attained a higher word frequency score overall in the COCA). Notably, Study 1 did not find a relationship between race and overall VF LF raw or scaled score, which suggests that this difference lays specifically within the quality of words produced rather than the number. Notably, however, this means that Black athletes likely performed better on this task, given that they demonstrated increased complexity of word production with equal number of words produced in a given amount of time. It is also important to consider that race may be a proxy for other variables. When taking a closer look at the race variable in this study, it is notable that the sample is unique and also includes many athletes who were born in countries other than Canada and/or spoke a language other than English as their first language. Therefore, before making conclusions about race generally, one must dig deeper to determine if there are other influential variables at play as well.

Overall, Study 2 did not replicate Abeare and Seguin's (2014) findings, as none of the word quality indexes developed were related to athlete performance on the TOPF. Study 1 also showed that there is no significant relationship between athlete performance on verbal fluency tasks and the TOPF.

The third study examined whether it was possible to use a unique combination of the variables and test scores to develop a regression-based approach to the estimation of premorbid functioning in this setting. It was hypothesized that the use of regression-based methods using multiple relevant predictor variables could improve the accuracy of premorbid functioning estimates beyond current approaches, while also emphasizing cultural competency in assessing diverse populations by incorporating a broader range of variables to create more precise and personalized estimates specific to this unique population.

Because Study 1 found mostly weak correlations between test scores and demographic, personal, and educational variables, developing regression equations with high predictive value was difficult. Therefore, the purpose of this study was updated to be focused more on better understanding these variables for the purpose of conducting better-informed neuropsychological assessment for this population. Regression analyses were completed for the ImPACT composite scores, Trail Making Test A and B, and D-KEFS Verbal Fluency subtasks. Overall, the TMT- B was found to have the best fit given the included predictor variables accounted for 37% of the variance for TMT-B. However, overall, there was evidence of only weak predictive power of the included demographic, personal, and educational variables in determining test performance.

Overall, there are several limitations to consider in the context of the findings of this dissertation. First, is that a wide range of factors can impact one’s cognitive performance across tests. This is demonstrated by the sheer number of correlations found between personal and demographic variables and test scores. As such, it strengthens the argument that no one variable or test score should be used to estimate premorbid functioning for the purpose of return-to-play decision-making. External influences can also impact test performance and were not included in these analyses. More specifically, things like physiological factors such as a poor night’s sleep, psychological factors such as a significant life stressor or depressive symptomatology, or performance-based anxiety were not included. As well, some of the included factors such as education are much more complex than they appear to be, just based on the "years of education" metric. For example, measuring educational quality is very complicated and encompasses interrelated factors such as achievement, school-specific features, and regional and local SES (Lezak et al. 2012).

Another salient limitation to the current study was that although the total sample size was sufficiently large, the sample size within groups, especially minority groups, often contained too few athletes to be able to make fair statistical conclusions. Instead, relationships were assessed for the purposes of exploration and providing future direction.

This study was also focused specifically on an athlete population, which was primarily White, born in Canada, English-speaking, a part of the men's football team, and enrolled in kinesiology. As such, this affects generalizability to other populations whose compositions are not consistent with this one (e.g., athletes from different sporting programs, age groups, and level of play, assessment of non-athlete populations who have sustained a concussion, and assessment with more general acquired brain injury populations or those with neurodegenerative conditions) and does not represent culturally diverse athletes' experiences very well. It is important to note, however, that this study was designed with athletes in mind, and so generalizability to other populations was not one of the goals. In addition, the sample is representative of the type and number of athletes that are typically referred for baseline testing by a campus sport concussion neuropsychology service.

Additionally, athletes tended to perform broadly in the Average range on the TOPF (most scores were between SS = 96 and 113.25, with the median being 105. Only 2.5% of scores were below SS = 85), and thus understanding relationships with other variables and the TOPF was difficult due to range restriction. Therefore, a greater and more representative sample size is needed to be able to generalize to other varsity athletics settings that do not mimic this sample's composition and to draw conclusions about differences in group performance for culturally diverse groups with very few athletes represented in this sample. This would include a greater representation of athletes who do not identify as White, primarily English-speaking, playing football, and born in Canada. It is important to note, however, that a restricted range was present for age and years of education, which are two variables that are often associated with cognitive performance in past research. Because this sample contained a restricted range for age and years of education, this likely impacted the findings as well.

Further, a limitation to verbally based tasks in general, is that they were conducted in English and there were at least 13 athletes whose first language was not English or English and another language. For example, a strong predictor of verbal fluency in English is proficiency in English, especially lexical knowledge, and so athletes with a greater lexical knowledge in English have been shown to generate more correct responses on verbal fluency tasks (Paap et al., 2019). As mentioned at the beginning of this dissertation, Rivera-Mindt (2010) and the AACN (2016, 2020) have challenged the discipline of neuropsychology to improve access to competent neuropsychological services to minoritized patients by increasing cultural competence through research. This dissertation aimed to better understand the influence of personal, educational, and demographic factors on cognitive test performance, which was accomplished. However, it is notable that several minoritized groups were not adequately represented in this sample due to small sample sizes.

Regarding future research, several different sociodemographic variables were identified as having a significant relationship with test performance, and although largely weak, their influence could be investigated further. In particular, the influence of first language and race on these tests should be further understood, given that there was a significant difference in terms of test performance on some verbal fluency tasks but a small sample size. Neuropsychologists working in sport neuropsychology may consider asking about grade 7/8 school performance, as self-reported academic performance at that timepoint seemed to correlate with several different test performances. As well, other personal variables such as parent education, enrollment in full vs. part-time studies, possession of a scholarship, and GPA were identified as being related to test performance.

Conducting studies that are more longitudinal in nature could also be helpful, whereby athletes' neuropsychological test performance and functioning post-injury is compared to their baseline test performance, and accuracy of estimation of premorbid functioning is evaluated across cognitive domains. More specifically, generating an estimate of an athlete's premorbid functioning using the findings from this dissertation and comparing that estimate with direct baseline test data for that athlete to determine how accurate the estimation was. This endeavor would be useful for optimizing the accuracy of estimation of premorbid functioning within the SRCC when baseline data is unavailable, as well as informing best practice approaches in all sport neuropsychology settings.

In addition, these findings could inform future research in the area of Artificial Intelligence (AI). It is likely that AI could be trained to estimate premorbid functioning based on a person's neuropsychological test scores, demographic variables, and personal variables post-concussion, as this type of task falls under the AI capability of predictive modeling and machine learning (Mollick, 2024). Once trained, the AI model could possibly predict premorbid functioning. It is important to note, however, that the accuracy of these predictions would depend on the quality and size of the dataset used for training, as well as the complexity of the variables being considered. Generally, more data is typically better for training AI models, especially when attempting to conduct a task as complex as estimating premorbid functioning. This dissertation has shown that this endeavor is quite complex, and that many different variables contribute to neuropsychologists' understanding of accurate estimation of premorbid functioning. It also emphasizes the importance of clinical judgement and a biopsychosocial approach to neuropsychological assessment. As such, the use of AI for this purpose is a possible future direction that could be informed by the relevant variables unveiled by this research project.

Lastly, the results of this dissertation highlight that baseline testing can be more effective than indirect approaches such as estimation, as long as athletes understand the significance of such testing and put forth their best effort. Therefore, a future direction should include the dissemination of these research findings into clinical practice by educating varsity athletes on the importance of baseline assessments for brain health post-concussion. To do this, a psychoeducational intervention could be designed and implemented to include content related to concussion education, the role of baseline testing, and the impact of effort on validity. Consultation with an interdisciplinary team should be pursued (e.g. athletes, athletic therapists, coaches, kinesiologists, and sport psychologists), thus promoting the development of educational resources for use in sport-related concussion centers, enhancing knowledge, safety, and overall health in sports.

In summary, the insights that were gleaned from this research project offer sport neuropsychologists a deeper understanding of the diverse demographic, educational, and personal factors that are important to consider when interpreting test performances. This study extended upon past research by encompassing a broader range of variables and examining them within the framework of baseline neuropsychological assessment. The identification of several different mildly influential variables ultimately underscores the significance of adopting a comprehensive biopsychosocial approach to neuropsychological assessment. As well, these findings not only contribute to the refinement of clinical practices but also highlights the importance of the consideration of the multifaceted nature of athletes' experiences when evaluating cognitive functioning.

With that said, these findings demonstrate that improving the estimation of premorbid functioning is a complex and challenging task. It remains uncertain whether any single method could achieve these goals with sufficient accuracy to be considered optimal and reliable, especially in return-to-play decision making. These and past approaches with other neuropsychological populations underscore the importance of a multifaceted and methodologically rigorous approach to improving the estimation of premorbid functioning in neuropsychology more generally, but also demonstrate that sport-related concussion may require a different approach than those used with other populations. In conclusion, the findings from this research do *not* inform an approach to estimating premorbid cognitive functioning that is better than baseline testing in sport neuropsychology, and thus removing baseline testing from sport-related concussion protocols in is not recommended.

REFERENCES/BIBLIOGRAPHY

AACN. (2016). *Relevance 2050 initiative. https://theaacn.org/relevance-2050/relevance-2050-initiative/*

AACN. (2021). *Position statement on use of race as a factor in neuropsychological test norming and performance prediction* [Position statement]. https://theaacn.org/wp-content/uploads/2021/11/AACN-Position-Statement-on-Race-Norms.pdf

Abeare, C., Erdodi, L., Messa, I., Terry, D. P., Panenka, W. J., Iverson, G. L., & Silverberg, N. D. (2021). Development of embedded performance validity indicators in the NIH Toolbox Cognitive Battery. *Psychological Assessment, 33*(1), 90–96. https://doi.org/10.1037/pas0000958

Abeare, C., Messa, I., Zuccato, B.G., Merker, B., & Erdodi, L. (2018). Prevalence of invalid performance on baseline testing for sport-related concussion by age and validity indicator. *JAMA Neurology, 75*(6), 697-703. doi:10.1001/jamaneurol.2018.0031

Abeare, C., & Seguin, A. (2014). *Predicting Estimated IQ with Verbal Fluency Response Characteristics.* Poster presented at the International Neuropsychological Society’s Annual Conference. Seattle, WA.

Acevedo, A., Loewenstein, D. A., Barker, W. W., Harwood, D. G., Luis, C., Bravo, M., Hurwitz, D. A., Aguero, H., Greenfield, L., & Duara, R. (2000). Category fluency test: Normative data for englishand spanish-speaking elderly.Journal of the International Neuropsychological Society, 6, 760– 769. https://doi.org/10.1017/S1355617700677032

Ackerman, P. L., & Heggestad, E. D. (1997). Intelligence, personality, and interests: Evidence for overlapping traits. *Psychological Bulletin*, *121*(2), 219–245. https://doi.org/10.1037/0033-2909.121.2.219

Adler, A. (1931). The practice and theory of individual psychology. Harcourt, Brace & Company.

Adler, A. (1937). Position in family constellation influences lifestyle. *International Journal of Individual Differences.* 3, 211-227.

Alves, L., Simões, M. R., & Martins, C. (2012). The estimation of premorbid intelligence levels among Portuguese speakers: the Irregular Word Reading Test (TeLPI). *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, *27*(1), 58–68. https://doi.org/10.1093/arclin/acr103

American Psychological Association. (2018). *Stress in America: Generation Z*. Retrieved from https://www.apa.org/news/press/releases/stress/2018/stress-gen-z.pdf

Ardila, A., Pineda, D., & Rosselli, M. (2000). Correlation between intelligence test scores and executive function measures. Archives of Clinical Neuropsychology, 15(1), 31–36. https://doi. org/10.1016/S0887-6177(98)00159-0

Arnett, J. J. (2000). Emerging adulthood: A theory of development from the late teens through the twenties. *American Psychologist, 55*(5), 469-480. https://doi.org/10.1037/0003-066X.55.5.469

Asken, B. M., Houck, Z. M., Clugston, J. R., Larrabee, G. J., Broglio, S. P., McCrea, M. A., McAllister, T. W., Bauer, R. M., & CARE Consortium Investigators (2020). Word-reading ability as a "hold test" in cognitively normal young adults with history of concussion and repetitive head impact exposure: A CARE Consortium Study. *The Clinical neuropsychologist*, *34*(5), 919–936. https://doi.org/10.1080/13854046.2019.1680735

Austin, M. P., Mitchell, P., & Goodwin, G. M. (2001). Cognitive deficits in depression: possible implications for functional neuropathology. *The British journal of psychiatry: the journal of mental science*, *178*, 200–206. https://doi.org/10.1192/bjp.178.3.200

Baddeley, A. D. (1986). Working memory. Oxford, England: Clarendon Press.

Baddeley, A. D. (2001). Is working memory still working? American Psychologist, 56, 851– 864

Balance Tracking Systems Inc. (2022). BTrackS assess balance system (Version 7.5) [Computer software].

Barker-Collo, S., Starkey, N., Lawes, C. M., Feigin, V., Senior, H., & Parag, V. (2012). Neuropsychological profiles of 5-year ischemic stroke survivors by Oxfordshire stroke classification and hemisphere of lesion. *Stroke*, *43*(1), 50–55. https://doi.org/10.1161/STROKEAHA.111.627182

Barona, A., Reynolds, C. R., & Chastain, R. (1984). A demographically based index of premorbid intelligence for the WAIS—R. *Journal of Consulting and Clinical Psychology*, *52*(5), 885–887. https://doi.org/10.1037/0022-006X.52.5.885

Barry, D., Bates, M. E., & Labouvie, E. (2008). FAS and CFL forms of verbal fluency differ in difficulty: A meta-analytic study. Applied Neuropsychology, 15, 97–106. https://doi.org/10.1080/ 09084280802083863

Beal, A. L. (1988). Canadian content in the WISC-R: Bias or jingoism. Canadian Journal of Behavioral Science, 20, 154–166Benton, A. L., Hamsher, K., & Sivan, A. B. (1994). Multilingual aphasia examination (3rd ed.). San Antonio, TX: The Psychological Corporation.

Berg, J. L., Durant, J., Banks, S. J., & Miller, J. B. (2016). Estimates of premorbid ability in a neurodegenerative disease clinic population: comparing the Test of Premorbid Functioning and the Wide Range Achievement Test, 4th Edition. *The Clinical Neuropsychologist*, *30*(4), 547–557. https://doi.org/10.1080/13854046.2016.1186224

Bishara, A. J., & Hittner, J. B. (2015). Reducing Bias and Error in the Correlation Coefficient Due to Nonnormality. *Educational and psychological measurement*, *75*(5), 785–804. https://doi.org/10.1177/0013164414557639

Blair, J.R., & Spreen, O. (1989). Predicting premorbid IQ: A revision of the National Adult Reading Test. *The Clinical Neuropsychologist, 3*, 129–136.

Broca, P. (1861). Remarques sur le siège de la faculté du langage articulé, suivies d’une observation d’aphémie (perte de la parole). *Bulletin de la Société Anatomique de Paris*, 6, 330–357.

Broglio, S. P., Cantu, R. C., Gioia, G. A., Guskiewicz, K. M., Kutcher, J., Palm, M., Valovich McLeod, T. C., & National Athletic Trainer's Association (2014). National Athletic Trainers' Association position statement: management of sport concussion. *Journal of athletic training*, *49*(2), 245–265. https://doi.org/10.4085/1062-6050-49.1.07

Brooks, B. L., Holdnack, J. A., & Iverson, G. L. (2011). Advanced clinical interpretation of the WAIS-IV and WMS-IV: Prevalence of low scores varies by level of intelligence and years of education. *Assessment, 18,* 156–167

Bryan, M. A., Rowhani-Rahbar, A., Comstock, R. D., Rivara, F., & Seattle Sports Concussion Research Collaborative (2016). Sports- and Recreation-Related Concussions in US Youth. *Pediatrics*, *138*(1), e20154635. https://doi.org/10.1542/peds.2015-4635

Burin, D. I., Jorge, R. E., Arizaga, R. A., & Paulsen, J. S. (2000). Estimation of premorbid intelligence: the word accentuation test-- Buenos Aires version. *Journal of Clinical and Experimental Neuropsychology*, *22*(5), 677–685. https://doi.org/10.1076/1380-3395(200010)22:5;1-9;FT677

Cattell, R. B. (1941). Some theoretical issues in adult intelligence testing [Abstract]. *Psychological Bulletin, 38,* 592.

Cattell, R. B. (1943). The measurement of adult intelligence. *Psychological Bulletin, 40,* 153–193. doi: 10.1037/h0059973

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: a critical experiment. *Journal of Educational Psychology, 54,* 1–22. doi: 10.1037/h0046743

Cattell, R. B. (1971). *Abilities: Their structure, growth and action.* Boston, MA: Houghton Mifflin

Chaudhry, M., & Ready, R. (2012). Differential effects of test anxiety & stress on the WAIS-IV. *Journal of Young Investigators, 24,* 60–66.

Cave, S.N., Wright, M., & von Stumm, S. (2022). Change and stability in the association of parents' education with children's intelligence. *Intelligence, 90*. https://doi.org/10.1016/j.intell.2021.101597

Chen, Y., Ho, M., Chen, K., Hsu, C., & Ryu, S. (2009). Estimation of premorbid general fluid intelligence using traditional Chinese reading performance in Taiwanese samples. *Psychiatry and Clinical Neurosciences*, *63*(4), 500-507. https://doi.org/10.1111/j.1440-1819.2009.01970.x

Cheng, W., & Ickes, W. (2009). Conscientiousness and self-motivation as mutually compensatory predictors of university-level GPA. *Personality and Individual Differences, 47*(8), 817–822. https://doi.org/10.1016/j.paid.2009.06.029

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Clay, M.B., Glover, K.L., Lowe, D.T. (2014). Epidemiology of concussion in sport: A literature review. *Journal of Chiropractic Medicine, 12*(4), 230-251. doi: 10.1016/j.jcm.2012.11.005.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Cortese, M. J., & Khanna, M. M. (2008). Age of acquisition ratings for 3,000 monosyllabic words. Behavior Research Methods, 40(3), 791-794

Crawford, J. R., & Allan, K. M. (1997). Estimating premorbid WAIS-R IQ with demographic variables: Regression equations derived from a UK sample. *The Clinical Neuropsychologist*, *11*(2), 192-197. https://doi.org/10.1080/13854049708407050

Crawford, J. R., Millar, J., & Milne, A. B. (2001). Estimating premorbid IQ from demographic variables: a comparison of a regression equation vs. clinical judgement. *The British Journal of Clinical Psychology*, *40*(1), 97–105. https://doi.org/10.1348/014466501163517

Crawford, J. R., Moore, J. W., & Cameron, I. M. (1992). Verbal fluency: A NART-based equation for the estimation of premorbid performance. British Journal of Clinical Psychology, 31, 327–329. https://doi.org/10.1111/j.2044-8260.1992.tb00999.x

Cullum, C. M., & Lacritz, L. H. (2009). Neuropsychological assessment in dementia. In M. F. Weiner, & A. M. Lipton (Eds.), The American Psychiatric Publishing textbook of Alzheimer disease and other dementias (pp. 85–103). Arlington, VA: American Psychiatric Publishing.

Cunningham, J., Broglio, S. P., O'Grady, M., & Wilson, F. (2020). History of Sport-Related Concussion and Long-Term Clinical Cognitive Health Outcomes in Retired Athletes: A Systematic Review. *Journal of Athletic Training*, *55*(2), 132–158. https://doi.org/10.4085/1062-6050-297-18

Dancey C.P., Reidy J. Pearson Education; 2007. Statistics without Maths for Psychology.

Davis, A. S., Finch, W. H., Drapeau, C., Nogin, M., E Moss, L., & Moore, B. (2016). Predicting verbal fluency using Word Reading: Implications for premorbid functioning. *Applied neuropsychology. Adult*, *23*(6), 403–410. https://doi.org/10.1080/23279095.2016.1163262

Davis, A.S., Bernat, D.J. & Reynolds, C.R. (2018). Estimation of Premorbid Functioning in Pediatric Neuropsychology: Review and Recommendations. *Journal of Pediatric Neuropsychology, 4,* 49–62. https://doi.org/10.1007/s40817-018-0051-x

Del Ser, T., González-Montalvo, J. I., Martínez-Espinosa, S., Delgado-Villapalos, C., & Bermejo, F. (1997). Estimation of premorbid intelligence in Spanish people with the Word Accentuation Test and its application to the diagnosis of dementia. *Brain and Cognition*, *33*(3), 343–356. https://doi.org/10.1006/brcg.1997.0877

Duff, K., Chelune, G. J., & Dennett, K. (2011). Predicting estimates of premorbid memory functioning: validation in a dementia sample. *Archives of Clinical Neuropsychology*, *26*(8), 701-705. https://doi.org/10.1093/arclin/acr083

Duff, K., Dalley, B., Suhrie, K. R., & Hammers, D. B. (2019). Predicting premorbid scores on the repeatable battery for the assessment of neuropsychological status and their validation in an elderly sample. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, *34*(3), 395–402. https://doi.org/10.1093/arclin/acy050

Dursun, S. M., Robertson, H. A., Bird, D., Kutcher, D., & Kutcher, S. P. (2002). Effects of ageing on prefrontal temporal cortical network function in healthy volunteers as assessed by COWA: An exploratory survey. Progress in Neuro-Psychopharmacology and Biological Psychiatry, 26, 1007–1010. https://doi.org/10.1016/S0278-5846(01)00321-9

Erdal, K. (2012). Neuropsychological testing for sports-related concussions: how athletes can sandbag their baseline testing without detection. *Archives of Clinical Neuropsychology*, *27*(5), 473-479. https://doi.org/10.1093/arclin/acs050

Erdodi, L., Korcsog, K., Considine, C., Casey, J., Scoboria, A., & Abeare, C. (2021). Introducing the ImPACT-5: An Empirically Derived Multivariate Validity

Composite. *The Journal of head trauma rehabilitation*, *36*(2), 103–113. https://doi.org/10.1097/HTR.0000000000000576

Erdodi, L. A., Sabelli, A. G., An, K. Y., Hastings, M., McCoy, C., & Abeare, C. A. (2020). Introducing a Five-Variable Psychiatric Screener based on the Visual Analog Scale (V-5). *Psychology & Neuroscience*, *13*(2), 219–239. https://doi.org/10.1037/pne0000201

Erlanger, D., Saliba, E., Barth, J., Almquist, J., Webright, W., & Freeman, J. (2001). Monitoring resolution of postconcussion symptoms in athletes: Preliminary results of a web-based neuropsychological test protocol. *Journal of Athletic Training*, *36*(3), 280–287.

Erlanger, D., Feldman, D., Kutner, K., Kaushik, T., Kroger, H., Festa, J., Barth, J., Freeman, J., & Broshek, D. (2003). Development and validation of a web-based neuropsychological test protocol for sports-related return-to-play decision-making. *Archives of Clinical Neuropsychology*, *18*(3), 293–316.

Eysenck, M. W. (1992). Anxiety: The cognitive perspective. Hove, England: Erlbaum.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior Research Methods, 39(2), 175–191. https://doi.org/10.3758/BF03193146

Faul, M., & Coronado, V. (2015). Epidemiology of traumatic brain injury. *Handbook of clinical neurology*, *127*, 3–13. https://doi.org/10.1016/B978-0-444-52892-6.00001-5

Fazio, V. C., Lovell, M. R., Pardini, J. E., & Collins, M. W. (2007). The relation between post concussion symptoms and neurocognitive performance in concussed athletes. *NeuroRehabilitation, 22*, 207–216.

Field, A. (2013). *Discovering statistics using IBM SPSS Statistics.* SAGE Publications Ltd.

Field, M., Collins, M. W., Lovell, M. R., & Maroon, J. (2003). Does age play a role in recovery from sports-related concussion? A comparison of high school and collegiate athletes. *The Journal of Pediatrics*, *142*(5), 546–553. https://doi.org/10.1067/mpd.2003.190

Flynn, J. R. (2007). *What is intelligence? Beyond the Flynn effect.* Cambridge University Press. https://doi.org/10.1017/CBO9780511605253

Flynn, J. R. (2013). The “Flynn Effect” and Flynn's paradox. *Intelligence, 41*(6), 851–857. https://doi.org/10.1016/j.intell.2013.06.014

Frey, M. C., & Detterman, D. K. (2004). Scholastic Assessment or g? The relationship between the Scholastic Assessment Test and general cognitive ability. *Psychological science*, *15*(6), 373–378. https://doi.org/10.1111/j.0956-7976.2004.00687.x

Fujii, D. (2017). *Conducting a Culturally Informed Neuropsychological Evaluation.* American Psychological Association.

Garber, H. L. (1988). *The Milwaukee Project: Preventing Mental Retardation in Children At Risk.* American Association on Mental Retardation.

Gaudet, C. E., & Weyandt, L. L. (2017). Immediate Post-Concussion and Cognitive Testing (ImPACT): a systematic review of the prevalence and assessment of invalid performance. *The Clinical neuropsychologist*, *31*(1), 43–58. https://doi.org/10.1080/13854046.2016.12206

Gershon, R. C., Wagster, M. V., Hendrie, H. C., Fox, N. A., Cook, K. F., & Nowinski, C. J. (2013). NIH toolbox for assessment of neurological and behavioral function. *Neurology*, *80*(11 Suppl 3), S2–S6. https://doi.org/10.1212/WNL.0b013e3182872e5f

Gilhooly, K. J., & Logie, R. H. (1980). Age-of-acquisition, imagery, concreteness, familiarity, and ambiguity measures for 1,944 words. *Behavior Research Methods & Instrumentation*, *12*(4), 395–427. https://doi.org/10.3758/BF03201693

Goodglass, H., & Kaplan, E. (1983). Boston diagnostic aphasia examination (BDAE). Philadelphia, PA: Lea & Febiger.

Goodwin, G.M. (1997). Neuropsychological and neuroimaging evidence for the involvement of the frontal lobes in depression. *Journal of Psychopharmacology, 11*(2), 115-122.

Gordon, K. E., & Kuhle, S. (2022). Canadians Reporting Sport-Related Concussions: Increasing and Now Stabilizing. *Clinical Journal of Sport Medicine: Official Journal of the Canadian Academy of Sport Medicine*, *32*(3), 313–317. https://doi.org/10.1097/JSM.0000000000000888

Government of Canada. (2019). Tackling the Problem Head-On: Sports-Related Concussions in Canada / Report of the Standing Committee on Health. Retrieved from: https://publications.gc.ca/site/eng/9.874202/publication.html

Gratz, K. L., & Roemer, L. (2004). Multidimensional assessment of emotion regulation and dysregulation: Development, factor structure, and initial validation of the difficulties in emotion regulation scale. *Journal of Psychopathology and Behavioral Assessment, 26*(1), 41-54.

Green, R.E.A., Melo, B., Christensen, B., et al. (2008). Measuring premorbid IQ in traumatic brain injury: An examination of the validity of the Wechsler Test of Adult Reading (WTAR). *Journal of Clinical and Experimental Neuropsychology, 30*, 163–172

Harnett, M. A., Godfrey, H. P., & Knight, R. G. (2004). Regression equations for predicting premorbid performance on executive test measures by persons with traumatic brain injuries.New Zealand Journal of Psychology, 33(2), 78–87.

Heaton, R., Grant, I., & Matthews, C. (1991). *Comprehensive norms for an expanded Halstead-Reitan Battery: Demographic corrections, research findings, and clinical applications.* Psychological Assessment Resources, Odessa, FL

Heaton, R. K., Miller, S. W., Taylor, M. J., & Grant, I. (2004). *Revised comprehensive norms for an expanded Halstead-Reitan battery: Demographically adjusted neuropsychological norms for African American and Caucasian adults.* Lutz, FL: Psychological Assessment Resources.

Heaton, R.K., Ryan, L., & Grant, I. (2009). Demographic influences and use of demographically corrected norms in neuropsychological assessment. In I. Grant & K.M. Adams (Eds.), *Neuropsychological assessment of neuropsychiatric and neuromedical disorders* (3rd ed.). New York: Oxford University Press.

Heinrichs, R. W., & Zakzanis, K. K. (1998). Neurocognitive deficit in schizophrenia: a quantitative review of the evidence. *Neuropsychology, 12*(3), 426-445. https://doi.org/10.1037/0894-4105.12.3.426

Helmes, E. (1996). Use of the Barona method to predict premorbid intelligence in the elderly. *Clinical Neuropsychologist*, *10*(3), 255-261. https://doi.org/10.1080/13854049608406688

Henry, J. D., & Crawford, J. R. (2004). A meta-analytic review of verbal fluency performance following focal cortical lesions. Neuropsychology, 18, 284–295. https://doi.org/10.1037/0894- 4105.18.2.284

Heyanka, D. J., Holster, J. L., & Golden, C. J. (2013). Intraindividual neuropsychological test variability in healthy individuals with high average intelligence and educational attainment. International Journal of Neuroscience, 123 (8), 526–531.

Holdnack, H.A. (2001). *Wechsler Test of Adult Reading: WTAR.* San Antonio. The Psychological Corporation.

Houck, Z., Asken, B., Clugston, J., Perlstein, W., & Bauer, R. (2018). Socioeconomic Status and Race Outperform Concussion History and Sport Participation in Predicting Collegiate Athlete Baseline Neurocognitive Scores. *Journal of the International Neuropsychological Society : JINS*, *24*(1), 1–10. https://doi.org/10.1017/S1355617717000716

Huang, M. (2001). Cognitive abilities and the growth of high-IQ occupations. *Social Science Research*, *30*(4), https://doi.org/10.1006/ssre.2001.0710

Iverson, G. L. (2007). Predicting slow recovery from sport-related concussion: The new simple-complex distinction. *Clinical Journal of Sport Medicine*, *17*(1), 31–37. https://doi.org/10.1097/JSM.0b013e3180305e4d

Iverson, G. L, Collins, M. W., Lovell, M.R. (2007). Predicting recovery time from concussion in high school football players. *Journal of the International Neuropsychological Society*, *13,* 65.

Iverson, G. L., & Schatz, P. (2015). Advanced topics in neuropsychological assessment following sport-related concussion. *Brain injury*, *29*(2), 263–275. https://doi.org/10.3109/02699052.2014.965214

Jencks, C., Smith, M., Acland, H., Bane, M., Cohen, D., Gintis, H., Heyns, B., & Michelson, S. (2014). Inequality: A Reassessment of the Effect of Family and Schooling in America. In D. B. Grusky (Ed. 4). *Social Stratification.* Routledge.

Jenkins, J. M., McGowan, P., Knafo-Noam, A., & Young, G. (2015). Rethinking the family as a context of human development: The case of siblings and individual development. International Journal of Behavioral Development, 39(4), 327–337.

Jenkinson, T., Muncer, S., Wheeler, M., Brechin, D., & Evans, S. (2017). Estimating verbal fluency and naming ability from the test of premorbid functioning and demographic variables: Regression equations derived from a regional UK sample. *British Journal of Clinical Psychology*, *57*(2), 135-147. https://doi.org/10.1111/bjc.12166

Johnstone, B. & Wilhelm, K.L. (1996). The longitudinal stability of the WRAT-R reading subtest: Is it an appropriate estimate of premorbid intelligence? *Journal of the International Neuropsychological Society*, *2*, 282–285.

Joseph, A. C., Lippa, S. M., McNally, S. M., Garcia, K. M., Leary, J. B., Dsurney, J., & Chan, L. (2021). Estimating premorbid intelligence in persons with traumatic brain injury: an examination of the Test of Premorbid Functioning. *Applied*

Neuropsychology: Adult, 28(5), 535-543. https://doi.org/10.1080/23279095.2019.1661247

Juhasz, B. J., Lai, Y. H., & Woodcock, M. L. (2015). A database of 629 English compound words: ratings of familiarity, lexeme meaning dominance, semantic transparency, age of acquisition, imageability, and sensory experience. *Behavior research methods*, *47*(4), 1004–1019. https://doi.org/10.3758/s13428-014-0523-6

Kaplan, E., & Kramer, J.H. (2001). Delis-Kaplan Executive Function System (D–KEFS) [Databaserecord]. APA PsycTests. https://doi.org/10.1037/t15082-000

Karr, J. E., Areshenkoff, C. N., & Garcia-Barrera, M. A. (2014). The neuropsychological outcomes of concussion: a systematic review of meta-analyses on the cognitive sequelae of mild traumatic brain injury. *Neuropsychology*, *28*(3), 321–336. https://doi.org/10.1037/neu0000037

Kendler, K. S., Turkheimer, E., Ohlsson, H., Sundquist, K., & Sundquist, J. (2015). Family environment and the malleability of cognitive ability: A Swedish national home-reared and adopted-away cosibling control study. *Proceedings of the National Academy of Sciences, 112*(15), 4612-4617. https://doi.org/10.1073/pnas.1417106112

Knight, R. G., McMahon, J., Green, T. J., & Skeaff, C. M. (2006). Regression equations for predicting scores of persons over 65 on the Rey auditory verbal learning test, the mini-mental state examination, the trail making test and semantic fluency measures. *British Journal of Clinical Psychology*, *45,* 393–402. https://doi.org/10.1348/014466505X68032

Kolb, B., & Whishaw, I. Q. (2003). Fundamentals of Human Neuropsychology. Worth Publishers.

Kroshus, E., Garnett, B., Hawrilenko, M., Baugh, C. M., & Calzo, J. P. (2015). Concussion under-reporting and pressure from coaches, teammates, fans, and parents. *Social Science & Medicine, 134,* 66–75. doi:10.1016/j.socscimed.2015.04.011

Krull, K., Scott, J., & Sherer, M. (1995). Estimation of premorbid intelligence from combined performance and demographic variables. *The Clinical Neuropsychologist, 9,* 83-87.

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of- acquisition ratings for 30,000 English words. Behavior Research Methods, 44(4), 978–990.

Langeluddecke, P. M., & Lucas, S. K. (2004). Evaluation of Two Methods for Estimating Premorbid Intelligence on the WAIS-III in a Clinical Sample. *The Clinical Neuropsychologist, 18*(3), 423–432. https://doi.org/10.1080/1385404049052422

Lehrl, S., Triebig, G., & Fischer, B. (1995). Multiple choice vocabulary test MWT as a valid and short test to estimate premorbid intelligence. *Acta neurologica Scandinavica*, *91*(5), 335–345. https://doi.org/10.1111/j.1600-0404.1995.tb07018.x

Lehtinen, N., Kautto, A., & Renvall, K. (2023). Frequent native language use supports phonemic and semantic verbal fluency in L1 and L2: An extended analysis of verbal fluency task performance in an L1 language attrition population. International Journal of Bilingualism, 0(0), https://doi.org/10.1177/13670069231193727

Leritz, E. C., McGlinchey, R. E., Lundgren, K., Grande, L. J., & Milberg, W. P. (2008). Using lexical familiarity judgments to assess verbally mediated intelligence in aphasia. *Neuropsychology, 22*(6), 687–696. https://doi.org/10.1037/a0013319

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *The Behavioral and brain sciences*, *22*(1), 1–75. https://doi.org/10.1017/s0140525x99001776

Lezak, M. D., Howieson, D. B., Bigler, E. D., & Tranel, D. (2012). *Neuropsychological assessment* (Fifth ed.). New York: Oxford University Press.

Lovell, M. R., Collins, M. W., Iverson, G., Johnston, K. M., & Bradley, J. P. (2004). Grade 1 or "ding" concussions in high school athletes. *American Journal of Sports Medicine, 32*, 47–54.

Lovell, M. R. (2022). *ImPACT Administration and Interpretation Manual.* San Diego, CA: ImPACT Applications, Inc.

Lovibond, S.H. & Lovibond, P.F. (1995). *Manual for the Depression Anxiety Stress Scales*. *(2nd. Ed.)* Sydney: Psychology Foundation.

Lubrini, G., Periáñez, J. A., Laseca-Zaballa, G., Bernabéu-Brotons, E., & Ríos-Lago, M. (2022). Verbal Fluency Tasks: Influence of Age, Gender, and Education and Normative Data for the Spanish Native Adult Population. *Archives of clinical neuropsychology: the official journal of the National Academy of Neuropsychologists*, *37*(2), 365–375. https://doi.org/10.1093/arclin/acab056

Lynn, R., & Meisenberg, G. (2010). National IQs calculated and validated for 108 nations. *Intelligence, 38*(4), 353–360. https://doi.org/10.1016/j.intell.2010.04.007

Mackinnon, A., & Mulligan, R. (2005). Estimation de l'intelligence prémorbide chez les francophones [The estimation of premorbid intelligence levels in French speakers]. *L'Encéphale: Revue de psychiatrie clinique biologique et thérapeutique, 31*(1), 31–43. https://doi.org/10.1016/S0013-7006(05)82370-X

Mackintosh, N. J. (2011). *IQ and Human Intelligence*. Oxford University Press.

Maerlender, A, Flashman, L., Kessler, A., Kumbhani, S., Greenwald, R., Tosteson, T., & Mcallister, T. (2010). Examination of the Construct Validity of ImpactTM Computerized Test, Traditional, and Experimental Neuropsychological Measures. *The Clinical Neuropsychologist, 24*(8), 1309–1325

Manderino, L., & Gunstad, J. (2018). Collegiate Student Athletes With History of ADHD or Academic Difficulties Are More Likely to Produce an Invalid Protocol on Baseline ImPACT Testing. *Clinical Journal of Sport Medicine: Official Journal of the Canadian Academy of Sport Medicine*, *28*(2), 111–116. https://doi.org/10.1097/JSM.0000000000000433

Manly, J. J., Jacobs, D. M., Touradji, P., Small, S. A., & Stern, Y. (2002). Reading level attenuates differences in neuropsychological test performance between African American and White elders. *Journal of the International Neuropsychological Society: JINS*, *8*(3), 341–348. https://doi.org/10.1017/s1355617702813157

Maroon, J. C., Winkelman, R., Bost, J., Amos, A., Mathyssek, C., & Miele, V. (2015). Chronic Traumatic Encephalopathy in contact sports: a systematic review of all reported pathological cases. *PLOS ONE, 10*(6), e0130507, https://doi.org/10.1371/journal.pone.0117338

Matarazzo, J. D. (1972). *Wechsler's Measurement and Appraisal of Adult Intelligence* (5th ed.). Oxford University Press.

Mathias, J. L., Bowden, S. C., & Barrett-Woodbridge, M. (2007a). Accuracy of the Wechsler Test of Adult Reading (WTAR) and National Adult Reading Test (NART) when estimating IQ in a healthy Australian sample. *Australian Psychologist*, *42*(1), 49–56. https://doi.org/10.1080/00050060600827599

Mathias, J.L., Bowden, S.C., Bigler, E.D. & Rosenfeld, J.V. (2007b). Is performance on the Wechsler Test of Adult Reading affected by traumatic brain injury? *British Journal of Clinical Psychology*, *46*, 457–466.

Matsuoka, K., Uno, M., Kasai, K., Koyama, K., & Kim, Y. (2006). Estimation of premorbid IQ in individuals with Alzheimer's disease using Japanese ideographic script (Kanji) compound words: Japanese version of National Adult Reading Test. *Psychiatry and clinical neurosciences*, *60*(3), 332–339. https://doi.org/10.1111/j.1440-1819.2006.01510.x

Mayers, L. B., & Redick, T. S. (2012). Clinical utility of ImPACT assessment for postconcussion return-to-play counseling: psychometric issues. *Journal of clinical and experimental neuropsychology*, *34*(3), 235–242. https://doi.org/10.1080/13803395.2011.63065

McCrea, M., Guskiewicz, K., Randolph, C., Barr, W., Hammeke, T., Marshall, S., . . . Kelly, J. (2013). Incidence, Clinical Course, and Predictors of Prolonged Recovery Time Following Sport-Related Concussion in High School and College Athletes. *Journal of the International Neuropsychological Society*, *19*(1), 22-33. doi:10.1017/S1355617712000872

McCrea, M., Hammeke, T., Olsen, G., Leo, P., & Guskiewicz, K. M. (2004). Unreported concussion in high school football players: Implications for prevention. *Clinical Journal of Sport Medicine, 14*, 13–17.

McCrory, P., Meeuwisse, W., Dvorak, J., Aubry, M., Bailes, J., Broglio, S., et al. (2017). Consensus statement on concussion in sports – The 5th international conference on concussion in sport held in Berlin, October 2016. *British Journal of Sports Medicine, 51*, 838–847.

McDermott, L. M., & Ebmeier, K. P. (2009). A meta-analysis of depression severity and cognitive function. *Journal of affective disorders*, *119*(1-3), 1–8. https://doi.org/10.1016/j.jad.2009.04.022

McFarlane, J., Welch, J., & Rodgers, J. (2006). Severity of Alzheimer's disease and effect on premorbid measures of intelligence. *British Journal of Clinical Psychology, 45*(4), 453–463. https://doi.org/10.1348/014466505X71245

McGurn, B., Starr, J.M., Topfer, J.A., et al. (2004). Pronunciation of irregular words is preserved in dementia, validating premorbid IQ estimation. *Neurology, 62*, 1184–1186.

McFarlane, J., Welch, J., & Rodgers, J. (2006). Severity of Alzheimer’s disease and effect on premorbid measures of intelligence. *British Journal of Clinical Psychology, 45*, 453–463.

McKenna, P., & Warrington, E. K. (1980). Testing for nominal dysphasia. *Journal of Neurology, Neurosurgery, and Psychiatry, 43,* 781–788. https://doi.org/10.1136/jnnp.43.9.781

Meehan, W. P., 3rd, Mannix, R. C., Stracciolini, A., Elbin, R. J., & Collins, M. W. (2013). Symptom severity predicts prolonged recovery after sport-related concussion, but age and amnesia do not. *The Journal of Pediatrics*, *163*(3), 721–725. https://doi.org/10.1016/j.jpeds.2013.03.012

Melrose, R. J., Campa, O. M., Harwood, D. G., Osato, S., Mandelkern, M. A., & Sultzer, D. L. (2009). The neural correlates of naming and fluency deficits in Alzheimer's disease: An FDG-PET study. International Journal of Geriatric Psychiatry, 24, 885–893. https://doi.org/10.1002/gps.2229

Merriam-Webster. (2021). *Premorbid.* Merriam-Webster Dictionary. https://www.merriam-webster.com/medical/premorbid

Meschyan, G., & Hernandez, A. (2002). Age of acquisition and word frequency: determinants of object-naming speed and accuracy. *Memory & Cognition*, *30*(2), 262–269. https://doi.org/10.3758/bf03195287

Messa, I., Korcsog, K., & Abeare, C. (2022). An updated review of the prevalence of invalid performance on the Immediate Post-Concussion and Cognitive Testing (ImPACT). *The Clinical Neuropsychologist*, *36*(7), 1613-1636. doi: 10.1080/13854046.2020.1866676.

Miller, L. S., & Rohling, M. L. (2001). A statistical interpretive method for neuropsychological test data. *Neuropsychology Review*, *11*(3), 143–169. https://doi.org/10.1023/A:1016602708066

Mollick, E. (2024). *Co-Intelligence: Living and Working with AI.* Portfolio.

Moser, R. S., Glatts, C., & Schatz, P. (2012). Efficacy of immediate and delayed cognitive and physical rest for treatment of sports-related concussion. *The Journal of pediatrics*, *161*(5), 922–926. https://doi.org/10.1016/j.jpeds.2012.04.012

Moraes, A. L., Guimar~aes, L. S., Joanette, Y., Parente, M. A. M. P., Fonseca, R. P., & Almeida, R. M. M. (2013). Effect of aging, education, reading and writing, semantic processing and depression symptoms on verbal fluency. Psicologia: Reflexao e Cr ~ ıtica, 26, 680–690. https://doi.org/10. 1590/s0102-79722013000400008

Moran T. P. (2016). Anxiety and working memory capacity: A meta-analysis and narrative review. *Psychological bulletin*, *142*(8), 831–864. https://doi.org/10.1037/bul0000051

Morrison, C. M., Chappell, T. D., & Ellis, A. W. (1997). Age of Acquisition Norms for a Large Set of Object Names and Their Relation to Adult Estimates and Other Variables. *The Quarterly Journal of Experimental Psychology Section A*, *50*(3), 528-559. https://doi.org/10.1080/027249897392017

Mullen, C. M., & Fouty, H. E. (2014). Comparison of the WRAT4 reading subtest and the WTAR for estimating premorbid ability level. *Applied Neuropsychology: Adult*, *21*(1), 69–72. https://doi.org/10.1080/09084282.2012.727111

NCS Pearson Corporation. (2009). *Manual for the test of pre-morbid functioning (TOPF)*. San Antonio, TX: Author.

Neisser, U., Boodoo, G., Bouchard, T. J., Boykin, A. W., Brody, N., Ceci, S. J., … & Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist, 51*(2), 77-101. https://doi.org/10.1037/0003-066X.51.2.77

Nelson, H. E. (1982). *National adult reading test (NART). Test manual*. Windsor, England: NFER-Nelson.

Nelson, H.E. & Willison, J.R. (1991). *The National Adult Reading Test (NART): Test manual* (2nd ed.). Windsor, UK: NFER- Nelson.

NINDS. (2017). NINDS common data elements. Retrieved from www.commondataelements.ninds.nih.gov

Norton, K., Watt, S., Gow, B., & Crowe, S. F. (2016). Are tests of premorbid functioning subject to the Flynn effect? *Australian Psychologist*, *51*(5), 374–379. https://doi.org/10.1111/ap.12235

Ollo, C., Lindquist, T., Alim, T.N., & Deutsch, S.I. (1995). Predicting premorbid functioning in crack-cocaine abusers. *Drug and Alcohol Dependence*, *40*, 173–175.

O'Rourke, J.J., Adams, W.H., Duff, K., et al. (2011). Estimating premorbid functioning in Huntington's disease: The relationship between disease progression and the Wide Range Achievement Test Reading subtest. *Archives of Clinical Neuropsychology, 26*, 59–66.

Paap, K. R., Mason, L. A., Zimiga, B. M., Ayala-Silva, Y., Frost, M. M., Gonzalez, M., & Primero, L. (2019). Other Language Proficiency Predicts Unique Variance in Verbal Fluency Not Accounted for Directly by Target Language Proficiency: Cross-Language Interference? *Brain sciences*, *9*(8), 175. https://doi.org/10.3390/brainsci9080175

Palmer, B. W., Appelbaum, M. I., & Heaton, R. K. (2004). Rohling's Interpretive Method and inherent limitations on the flexibility of "flexible batteries". *Neuropsychology review*, *14*(3), 171–169. https://doi.org/10.1023/b:nerv.0000048183.29813.15

Pearson Clinical (2017). *TOPF (test of pre-morbid function): Case studies*. San Antonio, TX: Pearson Clinical.

Pekkala, S., Goral, M., Hyun, J., Obler, L. K., Erkinjuntti, T., & Albert, M. L. (2009). Semantic verbal fluency in two contrasting languages. *Clinical linguistics & phonetics*, *23*(6), 431–445. https://doi.org/10.1080/02699200902839800

Piland, S. G., Ferrara, M. S., Macciocchi, S. N., Broglio, S. P., & Gould, T. E. (2010). Investigation of baseline self-report concussion symptom scores. *Journal of athletic training*, *45*(3), 273–278. https://doi.org/10.4085/1062-6050-45.3.273

Powell, B.D., Brossart, D.F., & Reynolds, C.R. (2003). Evaluation of the accuracy of two regression-based methods for estimating premorbid IQ. *Archives of Clinical Neuropsychology, 18*(3), 277-292. https://doi.org/10.1016/S0887-6177(02)00135-X

Psychological Corporation. (2001). *The Wechsler Test of Adult Reading (WTAR).* San Antonio, TX: Psychological Corporation.

Rabbitt, P., Mogapi, O., Scott, M., et al. (2007). Effects of global atrophy, white matter lesions, and cerebral blood flow on age- related changes in speed, memory, intelligence, vocabulary, and frontal function. *Neuropsychology*, *21*, 684–695.

Rabinowitz, A. R., & Arnett, P. A. (2012). Reading based IQ estimates and actual premorbid cognitive performance: discrepancies in a college athlete sample.

Journal of the International Neuropsychological Society, 18, 139-143. doi:10.1017/S1355617711001275

Rabinowitz, A. R., & Levin, H. S. (2014). Cognitive sequelae of traumatic brain injury. *The Psychiatric clinics of North America*, *37*(1), 1–11. https://doi.org/10.1016/j.psc.2013.11.004

Raven, J., & Raven, J. (2003). Raven Progressive Matrices. In R. S. McCallum (Ed.), *Handbook of nonverbal assessment* (pp. 223–237). Kluwer Academic/Plenum Publishers. https://doi.org/10.1007/978-1-4615-0153-4_11

Reitan, R. M. (1955). The relation of the trail making test to organic brain damage. *Journal of Consulting Psychology, 19,* 393-394.

Reynolds, C. (1997). Postscripts on premorbid ability estimation: Conceptual addenda and a few words on alternative and conditional approaches. *Archives of Clinical Neuropsychology*, *12*(8), 769-778. https://doi.org/10.1016/S0887-6177(97)00051-6

Reynolds, C. R., & Kaiser, S. M. (2003). Bias in assessment of aptitude. In C. R. Reynolds & R. W. Kamphaus (Eds.), *Handbook of psychological and educational assessment of children: Intelligence, aptitude, and achievement* (pp. 519–562). The Guilford Press.

Rivera Mindt, M., Byrd, D., Saez, P., & Manly, J. (2010). Increasing culturally competent neuropsychological services for ethnic minority populations: a call to action. *The Clinical neuropsychologist*, *24*(3), 429–453. https://doi.org/10.1080/13854040903058960

Russell, E. W. (1980). Fluid and crystallized intelligence: Effects of diffuse brain damage on the WAIS. *Perceptual and Motor Skills*, 51, 121–122.

Ryan, J. J., Paolo, A. M., & Dunn, G. E. (1995). Analysis of a WAIS R Old-Age Normative Sample in Terms of Gender, Years of Education, and Preretirement Occupation. *Assessment*, *2*(3), 225–231. https://doi.org/10.1177/1073191195002003003

Samuels, C., James, L., Lawson, D., & Meeuwisse, W. (2015). The Athlete Sleep Screening Questionnaire: a new tool for assessing and managing sleep in elite athletes. *British Journal of Sports Medicine, 418.* DOI: 10.1136/bjsports-2014-094332

Sandoval, T. C., Gollan, T. H., Ferreira, V. S., & Salmon, D. P. (2010). What causes the bilingual disadvantage in verbal fluency? The dual-task analogy. *Bilingualism: Language and Cognition*, *13*(2), 231–252. https://doi.org/10.1017/S1366728909990514

Sattler, J.M. (2001). *Assessment of children: Cognitive applications* (4th ed.). La Mesa, CA: Sattler.

Sattler, J.M. & Hoge, R.D. (2014). *Foundations of Behavioral, Social, and Clinical Assessment of Children* (6th ed.). Jerome M. Sattler, Publisher.

Saunders, R. L., & Harbaugh, R. E. (1984). The second impact in catastrophic contact-sports head trauma. *JAMA*, *252*(4), 538–539.

Schatz, P. (2010). Long-term test-retest reliability of baseline cognitive assessments using ImPACT. *American Orthopedic Society for Sports Medicine, 38*(1). https://doi.org/10.1177/0363546509343805

Schoenberg, M. R., Duff, K., Scott, J. G., Patton, D., & Adams, R. L. (2006). Prediction errors of the Oklahoma Premorbid Intelligence Estimate-3 (OPIE-3) stratified by 13 age groups. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, *21*(5), 469–475. https://doi.org/10.1016/j.acn.2006.06.006

Schretlen, D. J., Buffington, A. L. H., Meyer, S. M., & Pearlson, G. D. (2005). The use of word-reading to estimate "premorbid" ability in cognitive domains other than intelligence. *Journal of the International Neuropsychological Society, 11*(6), 784–787. https://doi.org/10.1017/S1355617705050939

Schretlen, D. J., Winicki, J. M., Meyer, S. M., Testa, S. M., Pearlson, G. D., & Gordon, B. (2009). Development, psychometric properties, and validity of the hopkins adult reading test (HART). *The Clinical Neuropsychologist*, *23*(6), 926–943. https://doi.org/10.1080/13854040802603684

Shipley, W.C. (1940). A self-administering scale for measuring intellectual impairment and deterioration. *Journal of Psychology, 9,* 371-377.

Scott, G. G., Keitel, A., Becirspahic, M., Yao, B., & Sereno, S. C. (2019). The Glasgow Norms: Ratings of 5,500 words on nine scales. *Behavior research methods*, *51*(3), 1258–1270. https://doi.org/10.3758/s13428-018-1099-3

Shao, Z., Janse, E., Visser, K., & Meyer, A. S. (2014). What do verbal fluency tasks measure? Predictors of verbal fluency performance in older adults. *Frontiers in psychology*, *5*, 772. https://doi.org/10.3389/fpsyg.2014.00772

Shura, R. D., Ord, A. S., Martindale, S. L., Miskey, H. M., & Taber, K. H. (2020). Test of Premorbid Functioning: You're Doing It Wrong, but Does It Matter?. *Archives of*

Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists, acaa025. Advance online publication. https://doi.org/10.1093/arclin/acaa025

Steward, K. A., Kennedy, R., Novack, T. A., Crowe, M., Marson, D. C., & Triebel, K. L. (2018). The Role of Cognitive Reserve in Recovery from Traumatic Brain Injury. *The Journal of head trauma rehabilitation*, *33*(1), E18–E27. https://doi.org/10.1097/HTR.0000000000000325

Storandt, M., Stone, K., & LaBarge, E. (1995). Deficits in reading performance in very mild dementia of the Alzheimer's type. *Neuropsychology*, *9*, 174–176.

Stosic, M. D., Murphy, B. A., Duong, F., Fultz, A. A., Harvey, S. E., & Bernieri, F. (2024). Careless Responding: Why Many Findings Are Spurious or Spuriously Inflated. *Advances in Methods and Practices in Psychological Science*, 7(1). https://doi.org/10.1177/25152459241231581

Strauss, E., Sherman, E., & Spreen, O. (2006). *A compendium of neuropsychological tests*: *Administration*, *norms, and commentary* (3rd ed.). New York: Oxford University Press.

Tallberg, I. M., Wenneborg, K., & Almkvist, O. (2006). Reading words with irregular decoding rules: a test of premorbid cognitive function?. *Scandinavian Journal of Psychology*, *47*(6), 531–539. https://doi.org/10.1111/j.1467-9450.2006.00547.x

Tamayo Martinez, N., Xerxa, Y., Law, J., Serdarevic, F., Jansen, P. W., & Tiemeier, H. (2022). Double advantage of parental education for child educational achievement: the role of parenting and child intelligence. *European journal of public health*, *32*(5), 690–695. https://doi.org/10.1093/eurpub/ckac044

Temple, R.O. (2019). The neuropsychology of concussion. In R.A. Carlstedt & M. Balconi (Eds. 1). *Handbook of Sport Neuroscience and Psychophysiology*, (360-373). Routledge.

Tsushima, W. T., Yamamoto, M. H., Ahn, H. J., Siu, A. M., Choi, S. Y., & Murata, N. M. (2021). Invalid Baseline Testing with ImPACT: Does Sandbagging Occur with High School Athletes?. *Applied Neuropsychology. Child*, *10*(3), 209–218. https://doi.org/10.1080/21622965.2019.1642202

Tyson, B. T., Shahein, A., Abeare, C. A., Baker, S. D., Kent, K., Roth, R. M., & Erdodi, L. A. (2023). Replicating a Meta-Analysis: The Search for the Optimal Word Choice Test Cutoff Continues. *Assessment*, *30*(8), 2476–2490. https://doi.org/10.1177/10731911221147043

van IJzendoorn, M. H., Juffer, F., & Poelhuis, C. W. K. (2005). Adoption and Cognitive Development: A Meta-Analytic Comparison of Adopted and Nonadopted Children's IQ and School Performance. *Psychological Bulletin, 131*(2), 301–316. https://doi.org/10.1037/0033-2909.131.2.301

van Kampen, D.A., Lovell, M.R., Pardini, J.E., Collins, M.W., & Fu, F.H. (2006). The "value added" of neurocognitive testing after sports-related concussion. *American Journal of Sports Medicine, 34*(10), 1630-1635.

Wechsler, David, 1896-1981. (1981). *WAIS-R : Wechsler adult intelligence scale-revised.* New York, N.Y. :Psychological Corporation

Wechsler, D. (1996). WISC-III manual Canadian supplement. Toronto: Harcourt Brace Canada.

Wechsler, D. (1997). *WAIS-III/WMS-III technical manual.* San Antonio, TX: Psychological Corporation.

Weschler, D. (2001). *Wechsler Adult Intelligence Scale-Third edition: Canadian technical manual.* Toronto, ON: Harcourt Canada.

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children-Fourth edition.* San Antonio, TX: Harcourt.

Wechsler, D. (2004). *Wechsler Intelligence Scale for Children–Fourth Edition: Canadian Technical Manual.* Toronto, Ontario, Canada: PsychCorp.

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale—Fourth Edition Administration and Scoring Manual.* San Antonio, TX: Pearson.

Wechsler, D. (2011). *The test of premorbid function (TOPF).* San Antonio, TX: The Psychological Corporation

Weickert, T., Goldberg, T.E., Gold, J.M., et al. (2000). Cognitive impairments in patients with Schizophrenia displaying preserved and compromised intellect. *Archives of General Psychiatry*, *57*, 907–913

Wernicke, C. (1874). *Der aphasische Symptomencomplex: Eine psychologische Studie auf anatomischer Basis*. M. Crohn und Weigert.

Wiens, A.N., Bryan, J.E., & Crossen, J.R. (1993). Estimating WAIS-R FSIQ from the National Adult Reading Test-Revised in normal subjects. *The Clinical Neuropsychologist*, *7*, 70–84.

Wilkinson, G. S. & Robertson G. J. (2017). *Wide Range Achievement Test, Fifth Edition* (WRAT5). Bloomington, MN: Pearson Inc.

Wilson, R. S., Rosenbaum, G., Brown, G., Rourke, D., Whitman, D., & Grisell, J. (1978). An index of premorbid intelligence. *Journal of Consulting and Clinical Psychology*, *46*(6), 1554–1555. https://doi.org/10.1037/0022-006X.46.6.1554

Wilkinson, G.S. (1993). *WRAT-3: The Wide Range Achievement Test administration manual* (3rd ed.). Wilmington, DE: Wide Range.

Wilkinson, G.S. & Robertson, G.J. (2006). *Wide Range Achievement Test 4 (WRAT4).* Lutz, FL: PAR (Psychological Assessment Resources).

Wilkinson, G.S. & Robertson, G.J. (2017). *Wide Range Achievement Test 5 (WRAT5).* Lutz, FL: PAR (Psychological Assessment Resources).

Zuccato, B.G. (2022). Factors affecting cognitive testing (FACT). *Unpublished Measure*.

APPENDICES

Appendix A: In-Person Baseline Test Battery

V8 Time 1
Factors Affecting Cognitive Testing (FACT) Questionnaire
Immediate Post-Concussion Assessment and Cognitive Testing (ImPACT)
- Word Memory
- Design Memory
- X's and O's
- Symbol Match
- Color Match
- Three Colors

National Institute of Health Toolbox Cognitive Battery (NIH-TB-CB)
- Picture Vocabulary Test
- Flanker Inhibitory Control & Attention Test
- List Sorting Working Memory Test
- Dimensional Change Card Sort Test
- Pattern Comparison Processing Speed
- Picture Sequence Memory
- Oral Reading Recognition Test
- FName Learning
- Auditory Verbal Learning Test
- Visual Reasoning Test
- Oral Symbol Digit Test
- FName Delay
- Auditory Verbal Learning Test (Delay)

Balance Testing and Neck Measurement
NIH-TB-CB Auditory Verbal Learning Test (Forced-Choice Recognition)
Word Choice Test
Test of Premorbid Functioning
Delis-Kaplan Executive Function System – Verbal Fluency Subtest (D-KEFS VF)
Trails A and B
V8 Time 2

Appendix B: Consent Form

SRCC Ongoing Consent Form

I agree to undergo a sports concussion evaluation at the direction of the Sports Related Concussion Centre (SRCC). I understand and agree that the results of this evaluation are to be the sole property of the SRCC. I hereby WAIVE ANY AND ALL CLAIMS that I have or may in the future have against the University of Windsor AND RELEASE the University of Windsor its officers, trustees, agents, and employees, from any and all liability for any loss, damage, expense or injury, including death, DUE TO ANY CAUSE WHATSOEVER, INCLUDING NEGLIGENCE, BREACH OF CONTRACT, OR BREACH OF ANY STATUTORY OR OTHER DUTY OF CARE ON THE PART OF SRCC.

I understand that the purpose(s) of this evaluation are for Sports Concussion baseline testing and, if deemed appropriate by the SRCC staff, a post-concussion evaluation. Should I be asked to undergo a post-concussion evaluation(s), I give permission for information to be provided by the Supervising Clinical Neuropsychologist(s) to the team Physicians and Athletic Therapists regarding my readiness for returning to play through one or more of the following means: face-to-face conversation, electronic communication, HeadCheck app, telephone or mailed letter. The Supervising Clinical Neuropsychologist also may discuss briefly the basis for this information in one or more of these communications with the sports medicine team (athletic trainers, team physicians, and personal physicians). I also consent to the sports medicine team sharing their observations of my behavior, the nature of my injury and my history with the SRCC.

Final return to play decisions will be made by the athlete in consultation with the Sports Medicine Team.

The post-concussion assessment may include an abbreviated clinical interview focusing on relevant background information. Otherwise, the baseline and post-concussion battery will generally consist of neuropsychological tests, balance testing, measures of socioemotional wellbeing, and a symptom surveys. Completion of this assessment typically requires 180 minutes.

Clinical and research evidence exists to suggest that this evaluation is useful in forming an opinion about readiness to return to play after sports concussion. Nonetheless, this evaluation does not provide all of the information that might be helpful in making such a decision or is it sufficient to provide a psychiatric diagnosis or a basis for academic, occupational, or rehabilitation planning or litigation. If you have elevated symptoms at baseline testing, we may contact you for a follow-up appointment in which we will attempt to determine the reason for the elevation at baseline.

I understand that I may withdraw my consent to this evaluation and to the transfer of information at any time by means of a written letter. If I do not withdraw my consent, it will remain effective.

I understand that I have the right to receive a copy of this form upon my request.

Confidentiality

The services that you will receive are confidential. This means that:

l. We will not give information to anyone outside of the Sports Medicine Team about your evaluation except as otherwise provided in this document.

2. Only in exceptional circumstances where required by law will members of the SRCC disclose information about you to others without your consent. For example, disclosure is required by law in cases where
(a) there is a suspicion of child abuse/neglect,
(b) there is suspicion of abuse/neglect of an elderly person in residential care,
(c) when a person poses a threat of serious injury to themselves or to others,
(d) abuse by a registered healthcare professional,
(e) subpoena of files by a court of law
(f) review by the College of Psychologists of Ontario (CPO)
(g) in the event of a public health emergency when contact tracing records are required.

Teaching

By indicating below, I give the SRCC Faculty and Staff permission to use my deidentified information collected pursuant to this document for classroom teaching or public workshops. To maintain confidentiality, I understand that my name, date of birth, or any other information that might identify me or my family will not be included in any presentations. I understand that giving or withholding permission will in no way affect the services that the SRCC is providing and that in the event that I give permission, I may rescind it at any time.
I consent to use of my **de-identified** (i.e., anonymous) data for teaching purposes.

- Yes
- No

Research

Information collected pursuant to this document will be included in a group database that could be used for studies to better our understanding of sports-related concussion. Information might also be included in a group database that could be used for studies to examine the utility and psychometric properties of our post-injury measures. Publications from this research would use deidentified data, with names, birthdates, and other personal information removed. No one can be identified in these publications, and your privacy is completely protected. All research to come from these data must be approved by the University of Windsor Research Ethics Board. Research conducted with collaborators outside of the SRCC will use deidentified data to maintain confidentiality. I give my consent to be contacted by the SRCC and/or SRCC members in the future for potential participation in research projects.

I consent to use of my **de-identified** (i.e., anonymous) data for research purposes.

- Yes
- No

I understand and consent to participating in the procedures involved in this assessment.

- Yes
- No

Signature:

Appendix C: Questionnaire Variables

Question	Response Options
Age	Text box
Date of Birth	Select month, day, year.
Sex at Birth	Male/Female
Gender Identity	Male/Female/Non-Binary or Third Gender/Two-Spirit/Prefer not to disclose
Transgender	Yes/No/Other - describe
Race/Ethnicity	Select all that apply: Hispanic, Latino, or Spanish origin, Indigenous North American, East Asian, West Asian, Black or African American/Caribbean, Pacific Islander, White, other, prefer not to disclose
Sport	Football/women's hockey/men's hockey/women's soccer/men's soccer/ women's basketball/men's basketball/women's volleyball/ men's volleyball/ women's track and field/ men's track and field/
Number of Years Playing Lancer sports	1-6+
Total Years Playing Sport	Text box
Highest Level of Education Completed	Drop down menu
Number of Times Participated in Baselines	0-6+
Handedness	Right, Left, Ambidextrous
Full-Time or Part-Time Studies	Full or Part-Time Studies
Year of University	1-6+
Current GPA	Text box
Screenshot of GPA	Upload
Grades 7-8 Average Marks in Math, Language Arts, Science, Social Studies, Art	ABCDF NA
Current Program of Study	All UWindsor program options + other
Have you Switched your Major?	Y/N
Previous degrees/diplomas	Yes – specify, no
Awards/Honours	Honour roll, OS, Board of Governor, distinction/great distinction, other
Accommodations	List of all accommodation options
Diagnosed or suspected LD	Reading, Writing, Math, Intellectual Disability, Autism, Language, Speech-Sound disorder, Tourette's Disease, ADHD, Other, prefer not to answer
Parent Highest Level of Education	Choices
Parent Occupation	Text box

Sibling Occupations	Text box
Position in Birth Order	1-4+ / prefer not to answer
Family Hx of LD, ADHD, ID, ASD, Trouble in school, other neurodevelopmental disorder	Choices
Parent Postal Codes	Text box
Knowledge of the following	Prenatal alcohol, cannabis, tobacco, other drugs, born premature, birth complications, medical illness, hospitalizations
First language	Choices
Ability to understand spoken English	Likert
Ability to understand written English	Likert
Ability to speak English	Likert
Ability to write English	Likert
English conversation skills	Choices
Birth country	Canada, other, prefer not to answer
Where did you grow up	Text
SES Ladder (best off/worst off)	Scale of 1-10
Family combined income	0-150K + choices
If not living with/dependent on family combined income	0-100K+
Employment status	Not working – employed 40+ hours per week
Current job	Text box
Hobbies	Text box
Past concussion Dx	Text box
Depression Anxiety Stress Scale (DASS)	Questionnaire
Difficulties in Emotion Regulation Scale (DERS)	Questionnaire
Athlete Sleep Screening Questionnaire (ASSQ)	Questionnaire
Patient-Reported Outcomes Measurement Information System (PROMIS) Measures	Questionnaire
Barkley Adult ADHD Rating Scale (BAARS-IV) Impulsivity Measures	Questionnaire

Appendix D: Description of ImPACT Subtests (Lovell, 2022)

Word Memory – The Word Memory subtest measures attention and verbal recognition memory. The examinee is presented with a list 12 words, twice, for 750 ms per word. They are then presented with a list of 24 words and asked to identify which words they had seen as part of the original list by clicking “yes” or “no” on the screen. Distractor words are chosen from the same semantic category as target words. Five versions of the word list are available to minimize practice effects. After a 20-minute delay (during which the examinee completes other subtests), the examinee is again asked to identify the words that were part of the original list.

Design Memory – The Design Memory subtest is designed to measure attention and visual recognition memory. The examinee is presented with a series of 12 designs, twice, for 750 ms per design. They are then presented with a series of 24 designs and asked to identify which designs they had seen before by clicking “yes” or “no” on the screen. Distractor designs are target designs that have been rotated in space. The designs were selected in order to make verbal encoding difficult, and different subsets of designs are available to reduce practice effects. After a 20-minute delay (during which the examinee completes other subtests), the individual is again asked to identify the designs they had seen as part of the group of designs.

X’s and O’s – The X’s and O’s subtest is designed to measure visual working memory and visual processing/visual motor speed. The examinee is presented with a distractor task, in which they are asked to press a specific key based on the image they see on the screen (e.g., “if you see a blue circle, press the “p” key on the keyboard”). After 74 completing the distractor task, they are presented with a screen of randomly assorted X’s and O’s which is displayed for 1.5 seconds. Each time the X’s and O’s are presented, three X’s or O’s are highlighted in yellow, and the examinee is asked to remember the location of the highlighted letters on the screen. Following the presentation of the letters, the distractor task is presented again to interfere with rehearsal. After completing the distractor task, the examinee is once again presented with a screen of X’s and O’s and asked to indicate which letters were previously highlighted. This process is repeated for four trials.

Symbol Match – The Symbol Match subtest is designed to measure visual processing speed, learning, and memory. The examinee is presented with a grid of the digits 1-9 paired with a common symbol. Symbols are readily identifiable (e.g., triangle, square, arrow). With the grid available to them, the examinee is presented with a symbol and asked to click, as quickly as possible, on the number that corresponds with that symbol. After 27 trials, the symbols from the grid are removed. The examinee is then again shown a series of symbols and asked to indicate, from memory, the number that was matched with each symbol.

Color Match – The Color Match subtest is designed to measure impulse control/response inhibition. The examinee is first asked to click a red, blue, or green button on the screen to ensure adequate color vision. After this, the examinee is presented with color words presented in a box in either the same color as the word, or in a different color (e.g., the word RED would be presented in red on color-congruent trials, and in another color on incongruent trials). The examinee is asked to click in the box as quickly as possible, but only if the word appears in the matching color.

Three Letters – The Three Letters subtest is designed to measure working memory and visual-motor response speed. The examinee is first presented with a distractor task, where they are presented with a randomly scattered grid of the numbers 1-25 and asked to count backwards from 25 by clicking on each successive number. Three consonants are then presented on the screen. The distractor task is then presented again for 18 seconds, after which the examinee is asked to recall the three letters by typing them on the keyboard. This process is repeated five times.

Appendix E: Athlete Major Area of Study, Race, First Language, Birth Country, and Mean Test Scores by Sport Type

	Football	Basketball		Hockey		Soccer		Volleyball	
	Male	Male	Female	Male	Female	Male	Female	Male	Female
Major									
Kinesiology	22%	13%	23%	0%	23%	-	38%	0%	14%
Business Administration	15%	25%	15%	50%	0%	-	0%	43%	8%
Criminology	11%	6%	0%	50%	8%	-	0%	0%	0%
Sport Management and Leadership	8%	12%	8%	0%	0%	-	0%	0%	8%
Psychology	7%	0%	0%	0%	15%	-	0%	0%	8%
Nursing	2%	6%	8%	0%	0%	-	12%	0%	0%
Other	35%	38%	46%	0%	54%	-	50%	57%	62%
Race									
White	62%	33%	58%	100%	92%	-	57%	72%	73%
Black	26%	46%	25%	0%	0%	-	14%	0%	9%
Asian	3%	14%	0%	0%	0%	-	29%	0%	0%
Hispanic	3%	0%	8%	0%	8%	-	0%	14%	9%
Indigenous	1%	0%	0%	0%	0%	-	0%	14%	0%
Other	5%	7%	8%	0%	0%	-	0%	0%	9%
First Language									
English	87%	88%	62%	100%	92%	-	86%	86%	100%
French	5%	12%	23%	0%	0%	-	14%	0%	0%
English and Arabic	1%	0%	0%	0%	0%	-	0%	0%	0%
English and French	2%	0%	8%	0%	0%	-	0%	0%	0%
English and Other	4%	0%	0%	0%	8%	-	0%	0%	0%
Other	1%	0%	8%	0%	0%	-	0%	14%	0%
Birth Country									
Canada	83%	79%	60%	100%	100%	-	100%	72%	91%
USA	9%	7%	10%	0%	0%	-	0%	14%	9%
Nigeria	1%	7%	0%	0%	0%	-	0%	0%	0%
Ivory Coast	0%	0%	10%	0%	0%	-	0%	0%	0%
Australia	0%	0%	10%	0%	0%	-	0%	0%	0%
Brazil	0%	0%	0%	0%	0%	-	0%	14%	0%
Cameroon	1%	0%	0%	0%	0%	-	0%	0%	0%
Dubai	1%	0%	0%	0%	0%	-	0%	0%	0%
Ethiopia	0%	7%	0%	0%	0%	-	0%	0%	0%
Turkey	1%	0%	0%	0%	0%	-	0%	0%	0%
Venezuela	0%	0%	10%	0%	0%	-	0%	0%	0%
Prefer Not To Answer	4%	0%	0%	0%	0%	-	0%	0%	0%
Mean Test Scores									
TOPF Raw	40.32	43.13	47.77	37.00	37.50	-	39.29	44.43	43.00

TOPF SS	102.88	104.93	111.00	98.00	99.83	-	103.14	108.14	106.18
ImPACT Verbal Memory	86.95	91.13	93.15	85.00	87.33	-	91.43	85.71	90.45
ImPACT Visual Memory	75.41	78.40	80.92	68.50	78.58	-	78.86	75.43	71.91
ImPACT Visual Motor Speed	39.40	42.25	42.59	40.14	44.37	-	39.84	41.26	43.87
ImPACT Reaction Time	0.61	0.60	0.63	0.61	0.55	-	0.58	0.59	0.56
ImPACT Impulse Control	5.28	4.07	5.15	2.50	4.08	-	5.57	2.57	5.64
TMT A Raw	23.99	25.47	21.69	24.00	21.75	-	20.71	22.43	20.45
TMT A T-Score	50.15	52.73	53.38	47.50	54.67	-	53.43	50.86	56.27
TMT B Raw	61.18	57.29	47.15	80.50	46.08	-	53.71	50.43	57.36
TMT B T-Score	50.14	53.43	56.08	38.00	55.67	-	53.71	58.29	52.55
D-KEFS VF-LF Raw	37.18	42.27	37.31	35.00	36.67	-	39.17	44.14	39.45
D-KEFS VF-LF Scaled Score	10.41	11.67	10.46	9.50	10.42	-	11.17	12.86	11.27
D-KEFS VF-CF Raw	44.83	48.33	42.00	34.00	43.83	-	45.50	46.86	51.27
D-KEFS VF-CF Scaled Score	12.77	14.40	11.85	8.50	12.50	-	13.17	13.71	15.09
D-KEFS VF-CS # Correct Raw	13.77	14.47	14.85	11.00	15.08	-	15.83	14.57	15.18
D-KEFS VF-CS # Correct Scaled Score	10.71	11.13	11.77	7.00	12.08	-	13.17	11.57	12.18
D-KEFS VF-CS Accuracy Raw	12.62	13.80	14.23	10.00	14.50	-	15.33	13.71	13.73
D-KEFS VF-CS Accuracy Scaled Score	11.15	11.87	12.46	8.00	12.92	-	14.00	12.29	12.36

Note. Major, race, first language, and birth country are presented as percentage of athletes by sport type. Test scores are displayed as average score by sport type. Because there was only one male soccer player, the "Men's Soccer" column has been removed for the purpose of deidentification and privacy.

VITA AUCTORIS

NAME:	Kassandra Helena Korcsog
PLACE OF BIRTH:	London, ON
YEAR OF BIRTH:	1995
EDUCATION:	Ste. Anne High School, Belle River, ON, 2013
	University of Windsor, B.Sc., Windsor, ON, 2017
	University of Windsor, M.Sc., Windsor, ON, 2018
	University of Windsor, M.A., Windsor, ON, 2020