A Statistical Examination of Impaired Performances Across Concussion Screening Instruments

Kathryn Ann Ritchie
Marquette University

Recommended Citation

Ritchie, Kathryn Ann, "A Statistical Examination of Impaired Performances Across Concussion Screening Instruments" (2017). *Master's Theses (2009 -)*. 403.
http://epublications.marquette.edu/theses_open/403

A STATISTICAL EXAMINATION OF IMPAIRED PERFORMANCES ACROSS CONCUSSION SCREENING INSTRUMENTS

by

Kathryn A. Ritchie, B. A.

A Thesis submitted to the Faculty of the Graduate School,
Marquette University,
in Partial Fulfillment of the Requirements for
the Degree of Master of Science

Milwaukee, Wisconsin

May 2017

ABSTRACT
A STATISTICAL EXAMINATION OF IMPAIRED PERFORMANCES ACROSS CONCUSSION SCREENING INSTRUMENTS

Kathryn A. Ritchie, B. A.

Marquette University, 2017

It is well documented that healthy individuals routinely obtain impaired scores on neuropsychological tests, which confounds the differential diagnosis process. Relatively little is known regarding the rates at which healthy individuals obtain impaired scores on measures that are used to detect cognitive symptoms associated with sports related concussion (SRC). The current study generated expected rates of impaired performance on the Standardized Assessment of Concussion (SAC), the Automated Neuropsychological Assessment Metrics Sports Battery (ANAM), Immediate Post-Concussion and Cognitive Testing (ImPACT), and Axon Sports (Axon) neurocognitive measures by conducting Monte Carlo analyses using data obtained from a large normative sample of amateur athletes. Consistent with a broad literature, approximately 20% of a non-injured sample would obtain at least one impaired score on these neurocognitive measures. Further, actual rates of impaired performance on the respective measures were investigated by stratifying an additional sample by estimated intellectual ability. Individuals with Above Average intellectual ability achieved impaired scores at a lower rate than individuals with Below Average intellectual functioning. This study elucidates the psychometric properties of commonly-used concussion screeners and should be considered when making return-to-play decisions.

ACKNOWLEDGEMENTS

Kathryn Ritchie, B. A.

I would like to thank my research mentor and committee chair, Dr. Jim Hoelzle, for his steadfast support and mentorship throughout this project. I would also like to thank Dr. Lindsay Nelson, not only for her guidance and input on the project as a committee member, but also for graciously allowing me to work with the data collected as part of Project Head-to-Head. I would also like to express gratitude to my other committee members, Dr. Stephen Saunders and Dr. Kristy Nielson for their contributions to the project. Additionally, I would like to thank members of the Neuropsychology and Personality Lab for their advice and encouragement in completing this project. I especially thank my family for their unyielding support and enthusiasm for my education. Finally, I would like to thank my friends and cohort members for their wholehearted support for my academic endeavors.

TABLE OF CONTENTS

LIST OF TABLES

Introduction

Though traumatic brain injury (TBI) is one of the most common health conditions in the United States, with prevalence estimates between 1.4 to 3 million cases per year (Summers, Ivins, & Schwab, 2009), its proper diagnosis, neurobiological mechanisms, and course remain somewhat elusive to medical professionals. Broadly, TBI is the result of brain injury occurring due to impact, acceleration, or deceleration (Lezak, Howieson, Bigler, & Tranel, 2012). *The Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* (*DSM-5*; American Psychiatric Association, 2013) defines TBI by the following constellation of symptoms: loss of consciousness (LOC), post-traumatic amnesia, or neurological indicators such as positive neuroimaging, new or markedly worse seizures, visual field deficits, olfaction impairment, or hemiparesis. Injury severity occurs on a spectrum, though it is often graded “mild,” “moderate,” or “severe” based on diagnostic factors, particularly a score on the Glasgow Coma Scale (GCS), a rapid assessment rating scale that quantifies neurological factors such as degree of consciousness, orientation, and the ability obey commands or respond to pain (Lezak et al., 2012).

The great majority (80%) of all TBI cases are considered mild in nature (Krauss, McArthur, Silverman, & Jayaraman, 1996). In an attempt to standardize the clinical conceptualization of mTBI, the American Congress of Rehabilitation Medicine Mild Traumatic Brain Injury Committee (1993) put forth guidelines for diagnosis. mTBI is conventionally defined as a head injury that is associated with one or more of the following symptoms: LOC for less than 30 minutes, any immediate retrograde or anterograde amnesia, or any disruption in mental state (i.e. confusion or disorientation) in conjunction with other neurological deficits (e.g., hemiparesis). Importantly, most experts agree that mTBI can occur without experiencing some of the characteristic symptoms, like LOC (Ruff et al., 2013). These diagnostic criteria are also consistent with the World Health Organization's Collaborate Task on Mild Traumatic Brain Injury (Holm, Cassidy, Carroll, & Borg, 2005), and were eventually adopted by the National Academy of Neuropsychology (Ruff, Iverson, Barth, Bush, & Broshek, 2009).

While males account for approximately two-thirds (59%) of all reported cases, mTBIs affect people of all ages, culture, and ethnicities (American Psychiatric Association, 2013; Summers et al., 2009). Additionally, mTBIs exhibit a unique, bimodal age distribution, occurring most frequently in childhood and then again in later adulthood (American Psychiatric Association, 2013; McCrea, 2008). This pattern indicates that mTBI affects individuals across the lifespan, and there are many unique developmental issues to consider when assessing for and managing symptoms associated with mTBI. Though their etiology is most frequently attributed to falls and motor vehicle accidents (MVAs) (American Psychiatric Association, 2013), it is often difficult to isolate confounding factors associated with these events. For example, investigation of MVAs often involves litigation-related issues and mTBI associated with falls are not typically observed and are significantly underreported (McCrea, 2008).

After sustaining mTBI, individuals may experience a constellation of transient neuropsychological and physiological symptoms (Gasonique, 1992). Physiologically, individuals may report dizziness, headache, fatigue, nausea, and balance problems. Common neuropsychological complaints include difficulties sustaining attention or concentration, memory problems, or confusion. These acute symptoms may make school or work more difficult in the days immediately following injury, but they often dissipate within days (McCrea, 2008). The majority of individuals fully recover from their symptoms within weeks to a few months post-injury (American Psychological Association, 2013).

Despite the fact that 90% of individuals recover from mTBI by six months post-injury (Roberts & Roberts, 2011), there exists a "miserable minority" of individuals who continue to experience residual post-concussive symptoms (PCS) outside of the normal recovery period. Various studies have estimated that this group may encompass 10-20% of all individuals who sustain mTBI (Ruff, Camenzuli, & Mueller, 1996; Ruff, 2005). There is some evidence that premorbid psychological factors including specific personality dimensions (i.e., alexithymia and anxiety sensitivity), somatization, low mood, and low levels of resiliency predict extended mTBI symptom experience (e.g., see McCauley et al., 2013; Nelson et al., 2016; Wood, O'Hagan, Williams, McCabe, & Chadwick, 2014;). Additionally, research has suggested that response bias and expectation may also affect how an individual experiences symptoms associated with mTBI (e.g., see Ferguson, Mittenberg, Barone, & Schneider, 1999; McCrea, 2008).

Systematic examinations of athletes who sustain sports-related concussion (SRC) have shed light on the recovery trajectory after an individual sustains mTBI. Concussion is a term frequently used in the athletic training literature, and is often considered synonymous with mTBI. After MVAs, sports injury is the second leading cause of concussion in adolescents and young adults (Sosin, Sniezek, & Thurman, 1996), comprising about 20% of all TBIs reported annually (Bailey, Barth, & McCrea, 2013).

Though a staggering 300,000 SRCs are reported each year, it is estimated that as many as half of all SRCs go unreported (McCrea, Hammeke, Olsen, Leo, & Guskiewicz, 2004). Due to improved awareness, knowledge, and societal concern, it is likely that the number of reported SRCs will increase in the future.

SRC provides a unique opportunity for prospective research. The first prospective study, Sports as a Laboratory Assessment Model (SLAM), collected and compared preseason baseline and post-concussion data to shed light on the acute effects of concussion and quantify recovery (Barth, Freeman, Broshek, & Varney, 2001; McCrea Broshek, & Barth, 2015). It is advantageous that SRCs are observed by others and are most likely to occur in young, healthy individuals who have motivation to recover quickly. In other words, many of the confounding factors associated with MVAs and unreported falls are not present in SRC (McCrea, 2008). A key finding from this literature is that athletes typically have a shorter recovery period than what is reported in the general population. McCrea and colleagues (2003) found that the majority of injured players experienced acute cognitive and balance symptoms *only* in the days following SRC. In particular, they observed a minor decline in scores on neuropsychological tasks measuring processing speed, learning, memory, and cognitive flexibility. On average, these scores returned to baseline levels within a week of sustaining SRC. While about 10% of the sample required more than a week to recover, none of the sample experienced any residual symptoms after 90 days post-injury. This influential model has shed light on the etiology, risk factors, and recovery process associated with SRC.

The prospective research model previously described has been widely applied by various research groups (e.g., see Belanger, Curtiss, & Vanderploeg, 2003; Echemendia et al. 2001) and adapted for evaluations in Emergency Departments (Sheedy, Geffen, Donnelly, & Faux, 2006). The extent to which the SRC model can be applied to nonathletes is debated. For example, it is clear that athletes are less likely to develop PCS symptoms than individuals with other etiologies (Bazarian et al., 1999). Additionally, researchers have proposed that there may be differences in the clinical presentation of individuals who sustain sports related and non-sports related injuries (Rabinowitz, Lei, & Levin, 2014). This is plausibly related to the fact that some cases of SRC may be associated with a lower degree of biomechanical force compared to other mTBIs (e.g. MVA), though it is readily acknowledged that athletes regularly encounter the biomechanical forces capable of causing mTBI (Guskiewicz, 2007). It has additionally been proposed that differences may be related to protective factors in athletes, which include both physiological (i.e. better-developed neck musculature) and psychological attributes (i.e. motivation to return-to-play, symptom underreporting, lower rates of psychopathology) (Rabinowitz, et al., 2014).

Despite the widespread prevalence of SRC, there is much ambiguity about how to quantify resulting cognitive impairments. SRCs are typically assessed first by on-site medical providers, including athletic trainers and/or team physicians. One problem associated with the "sideline" diagnosis of SRC is that proper medical personnel are not always available to assess SRCs, especially when they occur during practices or smaller events. As a result, the responsibility may fall on coaches, parents, teammates, and the individual athlete to accurately report and quantify symptoms associated with a concussion (Graham, Rivara, Ford, & Spicer, 2014). A well-documented and concerning factor that complicates sideline diagnosis is that some athletes underreport their symptoms. McCrea and colleagues (2004) surveyed athletes and found that 41% reported a hesitancy to report symptoms of SRC because they did not want to get taken out of the game, and 66% reported believing that their symptoms of SRC were not serious enough to necessitate medical attention.

Sideline evaluations of concussions are typically comprised of brief concussion screeners that assess the acute injury characteristics of SRC including neurological and neuropsychological status, symptom checklists, and balance ability. One of first and most widely-distributed cognitive measures in evaluating symptoms associated with mTBI is the Standardized Assessment of Concussion (SAC; McCrea, Kelly, & Randolph, 2000). The SAC was developed as a brief measure that integrated widely-utilized neuropsychological tasks sensitive to the neurocognitive symptoms associated with concussion. The measure is relatively easy to administer so that individuals who lack a background in assessment, such as athletic trainers or coaching personnel, could immediately evaluate an injured player (McCrea et al., 2000). Current practice involves administration of the SAC as part of the Sport Concussion Assessment Tool (SCAT3; McCrory et al., 2013), which contains items from the GCS, SAC, Balance Error Scoring System (BESS) and modified Maddocks questions that assess individuals' orientation to person, place, and time.

In the diagnosis of SRC, additional neuropsychological measures are also commonly administered during either sideline or baseline testing, or a hybrid of both (Iverson & Schatz, 2015). Often, neuropsychological evaluation of SRC is focused on the assessment of specific cognitive constructs including attention, processing speed, working memory, and executive functions (Randolph, McCrea, & Barr, 2005). The purpose of sideline testing is immediate, on-site assessment of injury severity, and is crucial in later making return-to-play decisions (Randolph et al., 2005). An athlete who performs at a level below expectation is assumed to be critically injured and will not return to competition. On the other hand, the primary goal of baseline testing is to track neurocognitive recovery status-post-injury by assessing return to an initial performance level. Typically, medical decisions are determined by evaluating whether a score is clinically impaired or meaningfully below a baseline level of performance. On most neuropsychological measures, scores are routinely considered “impaired” if they fall 1 to 1.5 standard deviations below an average level of performance. It is useful to have baseline testing for the sake of comparison because it controls for individuals’ premorbid levels of cognitive functioning. An athlete would not return to competition if his or her performance was well-below baseline performance, regardless of whether or not he or she scored in the impaired range.

Given the increased attention that mTBI has received in recent years, it is not surprising that a number of commercially available computerized neurocognitive test (CNT) batteries have been developed to evaluate symptoms associated with SRC. CNTs offer the convenience of testing multiple athletes at one time during the preseason, reduce possible practice effects, and diminish the human error of administration and scoring (Iverson & Schatz, 2015). They also offer the benefit of standardizing administration of the tests in the absence of a neuropsychologist, as tests can be administered by trainers or coaching personnel. Further, CNTs can measure reaction time more accurately, and offer many alternative testing forms for repeat testing (Collie, Darby, & Maruff, 2001). While there is currently no “gold standard” battery or measure of SRC in existence (Randolph, McCrea, & Barr, 2004), among the most widely-used CNTs are the Automated Neuropsychological Assessment Metrics Sports Battery (ANAM; ANAM Sports Medicine Battery, 2010), Axon Sports/Cogstate Sport (CogState, 2011), and Immediate Post-Concussion and Cognitive Testing (ImPACT, 2013).

Despite the promise of CNT batteries, an extensive 2005 literature review of computerized and traditional neuropsychological tests used to assess symptoms associated with SRC highlights significant psychometric limitations associated with these tools, which diminishes their clinical utility (Randolph et al., 2005). Specifically, in order for a test to be clinically useful, it should demonstrate adequate test-retest reliability, be able to detect cognitive symptoms associated with SRC (i.e., have adequate sensitivity), and discern cognitive impairment without behavioral symptoms. None of the batteries or measures investigated met each criterion. Given these shortcomings, the authors urged clinicians to use caution administering and interpreting performances on these measures.

Despite documented psychometric limitations, CNT batteries are becoming increasingly popular and widely utilized. It is estimated that about 40% of high schools that employ an athletic trainer use CNT batteries to aide return-to-play decision-making (Meehan, d’Hemecourt, Collins, Taylor, & Comstock, 2012). Weaknesses associated with these measures will be furthered expanded upon below to provide sufficient background and justify the current project.

Some of the limitations associated with CNT batteries are not unique per se to this group of tests. For example, many brief concussion measures have a low test ceiling, which diminishes the ability of a measure to detect cognitive symptoms associated with an injury. A ceiling effect occurs when test items are relatively easy and the majority of individuals achieve perfect scores, regardless of their injury status. For example, the SAC includes orientation questions (e.g., “What day of the week is it?”), that are so easy that they do not adequately distinguish between those who experience cognitive impairments due to concussion and those who do not, reducing the sensitivity of the measure overall. This limits the clinical utility of a measure, as perfect scores should represent excellent functioning rather than average performance (Brooks, Strauss, Sherman, & Iverson, 2009). Notably, a similar issue afflicts cognitive screeners used in other medical and psychiatric settings such as the Mini Mental Status Examination (MMSE; Folstein, Folstein, & McHugh, 1975).

An additional measurement issue that needs to be considered when utilizing and interpreting CNT batteries is that it is well documented that healthy individuals occasionally obtain low scores on neuropsychological measures. In contrast to the presence of ceiling effects, which can result in a false negative test score, characteristics of an individual and a specific test may interact and result in a false positive score (an impaired score in a normal functioning individual) (Brooks et al., 2009; Axelrod & Wall, 2007; Crawford, Garthwaite, & Gault, 2007). At the individual level, there is a greater range of natural variability in performance in “healthy” cognitive functioning than is often appreciated. For example, a review of the literature on abnormal neuropsychological scores in healthy individuals revealed that test batteries consisting of 20 measures typically yield at least two abnormal scores (Binder, Iverson, & Brooks, 2009). This phenomenon can be explained by measurement error associated with a test score and situational factors including fatigue, variable effort, and/or inattention (Binder et al., 2009). Additionally, an individual’s intellectual functioning will either increase or decrease the likelihood of a false positive score. On average, healthy individuals with above average intellectual functioning obtain fewer impaired scores than those with below average intellectual functioning (Brooks et al., 2009). That being said, however, it is not terribly uncommon for someone with high average abilities to obtain borderline impaired scores, especially on measures that are minimally correlated with general ability.

In addition to participant characteristics, the probability of observing impaired scores is also influenced by test factors. One relevant factor is the neuropsychological construct that the test aims to measure. For instance, tests that measure constructs that are less directly correlated to overall, general intelligence (e.g., processing speed) are more susceptible to yielding inaccurate scores (Donnell, Belanger, & Vanderploeg, 2011). Tests that are not highly correlated with general intelligence are more susceptible to regression towards the mean, which may under- or over-estimate true ability. Additionally, in a large-scale analysis of the psychometric properties of standardized neuropsychological batteries, researchers documented that neuropsychological tests with non-normalized distributions of scores (e.g., tests of verbal list-learning, executive functions, and visual memory) were more likely to yield erroneously abnormal scores (Donnell et al., 2011).

Another factor that may result in the presence of a false positive test score is the number of measures included in a test battery. For example, the likelihood of obtaining an impaired score increases based on the number of measures that are included in a test battery. Across studies, Binder and colleagues (2009) found that the median number of impaired scores was typically equal to 10-15% of the total test scores in each battery. For example, if the test battery included 30 measures, it would not be uncommon to observe three to five impaired performances. Using binomial probability modeling to approximate the frequency at which events occur in a normal distribution, Ingraham and Aiken (1996) found that with each additional test added to a battery, there was a greater probability of "normal" performance surpassing cutoff criteria for abnormal performance. The authors observed that when a sample completes a battery of six tests, at least 20 percent of the population is likely to obtain at least one impaired score.

In the previously described study, Ingraham and Aiken (1996) defined impairment as performance lower than one standard deviation below the mean. Logically, as more conservative cutoffs are selected to define impairment (e.g., lower than 1.5 standard deviations below the mean), the rate of observed low scores decreases (i.e., the presence of potentially false positive scores would decrease). The definition of impairment is a significant issue to consider when interpreting all test data. Requiring a more extreme score to define impairment will likely reduce the sensitivity of a measure, meaning that some injured individuals' neurocognitive symptoms will not be interrupted as problematic (Brooks et al., 2009). On the other hand, requiring a more extreme score to define impairment would decrease the likelihood of potentially misclassifying a healthy individual (i.e., specificity). Thus, there is a delicate balance that must be maintained between sensitivity and specificity when determining cutoff scores, especially for CNTs because many individuals administering and interpreting scores have a limited understanding of psychometric principles.

In summary, while it is relatively common for a healthy individual to obtain an impaired score on a neuropsychological measure, this fact presents a unique obstacle for

SRC assessment. Measures of SRC are primarily used to detect the presence of mTBI based on acute neuropsychological symptoms. The results of these tests help determine return-to-play decisions, yet the presence of an abnormal score does not always indicate that an individual has sustained SRC (Binder et al., 2009). The clear challenge for clinicians is determining when an "impaired" score reflects true impairment, relative to the normative sample or to a baseline performance. For example, injured athletes with above-average intellect may or may not display dramatic changes in neuropsychological functioning or score below an arbitrary cutoff for impairment. It is also difficult to determine whether a borderline score reflects normal variability in performance or neurocognitive symptoms associated with mTBI.

Current Study

Erroneous impaired scores affect clinical decision-making. If a healthy individual has an impaired score at baseline, it will confound return-to-play decisions. Conversely, if an individual who has fully recovered from SRC elicits an abnormally low score, the individual will not appear to be recovered. To facilitate interpretation of neuropsychological test performance in the context of SRC and mTBI in general, this research aims to investigate how frequently "impaired" scores are expected and observed in non-injured participants.

As previously mentioned, Ingraham and Aiken (1996) conducted early studies on the base rate of impaired neuropsychological scores using binomial probabilities. A significant issue with this methodological approach is that it assumes all measures are uncorrelated. This is problematic since many neuropsychological measures are at least moderately correlated with one another, and utilizing binomial probabilities on correlated measures is likely to inflate the probability of observing an impaired score (Crawford et al., 2007; Decker, Schneider, & Hale, 2012). A decade later, addressing this limitation, Crawford and colleagues developed a Monte Carlo simulation procedure to generate expected base rates of impaired performance on correlated tasks. Briefly, the correlation matrix describing relationships between tests is taken into account when generating one million simulated participant scores. After scores are generated, one can simply determine the likelihood of observing an impaired score (or any other score). An advantage of this approach is that it provides more accurate base rate information with correlated measures and is robust to abnormal score distributions (Crawford et al., 2007). As such, this method will be utilized to evaluate SRC measures.

The literature contains numerous examples of common neuropsychological measures, including the Wechsler Adult Intelligence Scale, Third Edition (WAIS-III; Wechsler, 1997), the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS; Randolph, 1998; Crawford, Garthwaite, Morris, & Duff, 2012) and the Delis-Kaplan Executive Function System (D-KEFS; Delis, Kaplan, & Kramer, 2001; Crawford, Garthwaite, Sutherland, & Borland, 2011), that have been investigated using both binominal probability and the Monte Carlo simulation approaches. This vast body of literature makes clear that normal individuals commonly obtain lower than expected scores. For example, when Crawford and colleagues (2007) first applied this method to the WAIS-III, they observed that over 34% of the population was expected to exhibit at least one Index score lower than a Scaled Score of 85 (equal to or lower than one standard deviation below the mean).

Crawford and colleagues' (2007) procedures have also been employed in evaluating base rates of clinically significant scores on personality measures. Investigation of the Minnesota Multiphasic Personality Inventory, Second Edition (MMPI-2; Butcher, 2001) suggested that 36.8% of normal individuals were expected to obtain clinically significant scores, defined as 65T or above (Odland, Martin, Perle, Simco, & Mittenberg, 2011). As a whole, this research indicates that over a third of individuals are expected to obtain clinically significant scores on commonly administered measures such as the MMPI-2 or WAIS-III, implying that it is relatively common to observe an "impaired" or "clinically significant" score in normative samples. This feature of psychological instruments does not imply that all measurement tools are flawed, but it does indicate that psychometric scores are more variably distributed than one may assume, meaning that it is important to consider factors beyond isolated scores during the assessment process.

In the only published application of Monte Carlo analyses to SRC batteries to date, Nelson (2015) used simulations to estimate the base rate of individuals classified as impaired given various reliable change index (RCI) thresholds. Consistent with previous literature, Nelson observed that a meaningful percentage healthy individuals would be classified as impaired based on RCI cutoffs, test battery size, test battery intercorrelations, and the criteria applied to define "impairment." Specifically, as more RCIs were interpreted and the threshold for impairment was more liberal, the frequency of individuals classified as impaired increased (Nelson, 2015).

Results from Monte Carlo analyses across diverse measures consistently document that it is not uncommon for healthy individuals to obtain scores at or below the 5th percentile (corresponding with scores approximately 1.5 standard deviations below the mean). This body of literature also suggests the likelihood of a healthy individual obtaining an impaired score increases as more scores are interpreted. These two important aspects of previous research—insufficient psychometric properties of concussion screeners, as well as the evidence that healthy individuals sometimes obtain clinically significant scores—provide the need and justification for a statistical investigation of instruments used to evaluate symptoms associated with SRC to document how often uninjured individuals obtain impaired scores. While Monte Carlo analyses have been applied to other commonly used psychological assessment tools to determine the rate at which impaired scores naturally occur (Crawford et al., 2007), there has been limited research to date on the frequency of obtaining an impaired score on measures of SRC.

The primary aim of the current study is to generate expected frequencies of impaired scores on both traditional and computerized measures of SRC. To achieve this goal, Monte Carlo analyses will be conducted using correlation matrices, generated from a large community sample of healthy amateur athletes, following Crawford and colleagues' (2007) method. This study will be the first to present simulated base rates of commonly used measures of SRC, and will permit for a comparison between measures. Ultimately, this aim should improve clinical practice; if a high frequency of impaired scores is expected in healthy participants on a certain measure, this suggests relatively limited clinical utility in identifying cognitive symptoms in the post-acute stage of mTBI.

It is hypothesized that Monte Carlo analyses will suggest that healthy individuals are likely to obtain impaired scores on the SAC, ANAM, ImPACT, and Axon, defined by performance below the fifth percentile. It is expected that this rate will approach 20% based on empirical investigation of other neuropsychological measures (Crawford et al., 2007). Further, it is anticipated that ANAM, Axon, and ImPACT will yield higher base rates of impaired scores in normal participants than the SAC. This hypothesis is supported by research documenting that these computerized batteries have variable reliability and validity (Broglio, Ferrara, Macciocchi, Baumgarter, & Elliot, 2007; Resch, et al., 2013). This suggests that scores derived from computerized measures include more error variance than scores derived from the SAC[1]. Greater error variance increases the likelihood that a normal participant might obtain an impaired score. Amidst the uncertain utility of SRC measures, a goal of this study is to determine which measure yields the lowest base rate of impaired scores in non-injured participants.

The second primary aim of the proposed study is to stratify an additional sample by participants' estimated intellectual functioning to examine whether actual rates of impaired scores vary among individuals of Below Average, Average, and Above Average intellectual functioning. It is hypothesized that the lowest base rate of impaired performances will be observed in the sample with above average intellectual functioning based on Brooks and colleagues' findings (2009). That being said, it should be noted that there are several compelling reasons to expect that non-injured athletes, regardless of intellectual ability, may have impaired scores. For example, results of baseline testing

---

[1] That is not to say that the SAC does not have error variance, but the latter appears related to a ceiling effect, which should decrease the likelihood of uninjured athletes obtaining impaired scores during baseline testing

may be used to determine return-to-play decisions. Thus, athletes may be motivated to purposely decrease performance at baseline so that a score obtained after injury more closely approximately the invalid baseline score.

Methods

Participant Samples

The current study made use of data collected from amateur athletes. The data were collected as part Project Head-to-Head, a large-scale study of the neuropsychological outcomes of concussion conducted through the Medical College of Wisconsin and U.S. Department of Defense's joint project. These data have previously been used to explore the reliability and validity of SRC assessment measures (Nelson et al., 2016a), investigate the rate of invalid baseline performances in a prospective athlete sample (Nelson, Pfaller, Rein, & McCrea, 2015), and examine the role of pre-injury somatization symptoms on SRC recovery (Nelson et al., 2016b).

The sample consists of 2,159 amateur athletes recruited from high schools and colleges in southeastern Wisconsin. The mean age of the sample is 17.78 (SD = 1.93), with an estimated mean IQ score of 101.40 (SD = 12.57). Over two-thirds of the sample is male (77.0%) and white (80.5%). The athletes sampled are involved in a number of different contact sports, with most participation in men's football (49.1%). Additionally, 4.8% of the current sample sustained a concussion while participating in contact sports during the season after baseline testing was completed. Athletes completed numerous traditional and computerized neurocognitive measures, balance testing, and symptom self-report questionnaires. Data collection was such that each participant completed the SAC and two of the three CNTs that were examined.

Procedure

Data collection was approved by the Medical College of Wisconsin's Institutional Review Board. Amateur athletes were recruited from Milwaukee-area high schools and colleges. Each participant went through an informed consent process. Minors under the age of 18 received parent or guardian consent and assent prior to participating. All participant identifying information was stored separately from their testing data, which was coded using a generic identification number and stored in Research Electronic Data Capture (REDCap, Harris et al., 2009), a secure web-based data storage platform for research. Only principal and co-investigators have access to the data.

Of the total sample (*n* = 2,159), 37 (1.7%) were excluded for invalid performance on the Medical Symptom Validity Test (Green, 2004). The remaining sample was randomly divided into two groups. Data from one group were used to generate correlation matrices for each measure of interest. Correlation matrices are required to conduct Monte Carlo simulations. This is a necessary step because correlation matrices have not been published from sufficiently large samples. The other group was stratified by estimated intelligence level and the frequency of impaired scores was investigated by group. Wechsler Test of Adult Reading (WTAR; Wechsler, 2001) scores were used to establish three groups: Below Average, Average, and Above Average intellectual ability. Standard scores below 90 were considered Below Average; scores of 90-109 were considered Average, and those 110 or above were classified as Above Average.

Measures

Standardized Assessment of Concussion (SAC). Conceptually, the SAC consists of four separate content areas that contribute to an overall score: orientation, immediate memory, concentration, and delayed memory. The test was initially normed with a sample of athletes, ranging from 10-25 years old, but extended norms are available for both younger and older age groups. While test-retest studies have ranged from .31 to .71, researchers determined the overall reliability of the SAC was .64 across all studies (McCrea et al., 2000). Since the SAC is comprised of iterations of other neuropsychological tests, including a brief digit-span working memory task and a short list-learning and recall task (see Appendix A), McCrea and colleagues argued that the test has sufficient content validity. Additionally, with respect to group classification, Barr and McCrea (2001) found that a 1-point reduction on the SAC score at retesting accurately categorized injured and non-injured athletes with 94% sensitivity and 76% specificity.

Immediate Post-Concussion and Cognitive Testing (ImPACT). ImPACT is a CNT consisting of six tasks that contribute to Verbal Memory, Visual Memory, Visual Motor Speed, Reaction Time, and Impulse Control composite scores (see Appendix B). The Impulse Control score is an embedded measure of effort and performance validity. ImPACT has been found to yield variable psychometric properties, which may contribute limited clinical utility. Nelson and colleagues (2016a) found a seven day test-retest reliability of .61 and sensitivities to SRC across composite scales varied from 24.4 to 39.5%.

Axon Sports/CogState Sports (Axon). Axon is a computerized battery that consists of four neurocognitive tasks that measure processing speed, attention, visual memory, and working memory (see Appendix C). In their review of CNT batteries, Resch and colleagues (2013) documented variable psychometric properties on this measure. Reliability coefficients ranged from .45 to .90, though there was some evidence of criterion validity when compared to traditional neuropsychological tests, with correlation coefficients varying between .23 and .83, across subtests. Due to the great variation in these coefficients, Resch and colleagues concluded that the psychometric properties of this battery are less than ideal and questioned whether the measure had sufficient clinical utility. In a prospective SRC study, Axon's seven day test-retest reliability was .60, and it yielded a sensitivity to concussion across subtests of 6.8 to 48.6% (Nelson et al., 2016a). These findings suggest that Axon also has variable psychometric properties in an athlete sample.

Automated Neuropsychological Assessment Metrics Sports Battery (ANAM). ANAM is a battery of 12 common computerized neuropsychological tasks (i.e. simple reaction time and spatial processing) with an administration time of approximately 30 minutes (see Appendix D). Resch and colleagues' (2013) extensive review indicated that ANAM subtests have test-reliability coefficients ranging from .14 to .86, depending on the length of time between testing sessions Additionally, it was reported that correlations among subtests vary between -.01 to .65. There is weak discriminant validity between ANAM working memory subtests and other validated measures of sustained attention.

The ability of ANAM subtest performance to identify concussed athletes varied between 1% to 15% (specificity ranged from 86% and 100%; Resch, McCrea, & Cullum, 2013). In a large scale, prospective SRC study, seven day test-retest reliability was .65, while its sensitivity to concussion ranged from 6.0 to 23.8% across subtests (Nelson et al., 2016a).

Data Analyses

All statistical analyses were performed using SPSS software, Version 24 (SPSS, Inc., Chicago, IL) and Crawford's (2007) macro to conduct Monte Carlo simulations. To address the first primary aim, Monte Carlo simulations were conducted. Crawford's (2007) program was utilized to estimate percentages of impaired scores in a battery (PercentAbnormK.exe; http://homepages.abdn.ac.uk/j.crawford/pages/dept/psychom.htm). Briefly, to conduct Monte Carlo simulations, a Cholesky decomposition (the square root of R) is derived from the correlation matrix (R), describing associations between subtests/indices for a respective measure. A Cholesky decomposition is a mathematical process used to define a lower triangular matrix. In a lower triangular matrix the entries above the diagonal in a square matrix are zero. Next, one million random vectors (i.e., normally distributed test scores), based on the number of measures in each battery, are generated and postmultiplied by the Cholesky decomposition. This process results in one million sample "scores," which are considered when determining the probability of obtaining any given score. In short, the Monte Carlo simulations made use of measure-specific correlation matrices to determine how frequently certain scores occur in the simulated distribution.

To address the second aim, a series of one-way ANOVAs were completed to examine mean-level group differences across measures. The Bonferroni correction was applied to account for multiple comparisons. Welch tests and Games-Howell tests were used when equal variances were not assumed among groups. For each measure, actual frequencies of scores below the 5th percentile were also computed, and a series Chi-square test were conducted to determine whether rates of obtaining one or more impaired scores varied among estimated IQ groups. To examine group differences, post-hoc analyses of adjusted residuals, corrected for family-wise error were conducted (see Sharpe, 2015).

Results

General Overview. Analyses and results relevant to Aim 1 and 2 will be sequentially presented by instrument. Given that each participant did not complete all four measures, sample sizes vary by Aim and measure. For Aim 1, approximately 30% of the entire sample was randomly selected to generate a stable correlation matrix for each measure. This percentage was selected to ensure an adequate sample size of at least 300 (Crawford et al., 2007). The remaining data from the sample were used to address Aim 2 and intelligence groups were based on WTAR performance.

SAC

Aim 1. Bivariate Pearson correlations revealed limited associations among Index scores (see Table 1). Notably, a weak positive association was observed between SAC Orientation and Immediate Memory scores ($r = 0.14$). There was also a weak association between Immediate Memory and Concentration scores ($r = 0.21$). All Indices displayed at least weak associations with WTAR standard score ($r \geq 0.08$) except SAC Orientation, which was not significantly correlated ($r = 0.02$).

Table 1.

Correlations among SAC scores

Measure	1.	2.	3.	4.	5.
N	606	606	606	606	605
1. Orientation Total Score	-				
2. Immediate Memory Total Score	.14**	-			
3. Concentration Total Score	.02	.21**	-		
4. Delayed Recall Total Score	.04	.06	.08	-	
5. WTAR Standard Score	.02	.26**	.30**	.08*	-

Note: * indicates significance of $p < .05$, ** indicates significance of $p < .001$

Table 2 reports the results of Monte Carlo analyses utilizing different cutoffs to define an “impaired” score. When defining impaired scores as those falling below the lowest 5th percentile, 17.68% of healthy individuals would be expected to obtain at least one impaired SAC score, while 2.14% would be expected to obtain at least two impaired scores. Utilizing a more liberal criterion for impairment increases the expected frequency of impaired scores. For example, when using a 10th percentile cutoff, 32.89% of healthy individuals would be expected to obtain at least one impaired SAC score, while 6.32% at least two impaired scores. Defining “impaired” as the lowest 15% of the distribution yielded a base rate of obtaining at least one impaired score at 45.50%, with 12.32% expected to obtain at least two impaired scores.

Table 2.
Simulated frequency of impaired scores on the SAC

	Number of Impaired Scores Observed		
Cutoff	1	2	3
<5th Percentile	17.68%	2.14%	0.18%
<10th Percentile	32.89%	6.32%	0.74%
<15th Percentile	45.50%	12.32%	2.02%

Aim 2. One-way ANOVAs were completed to examine differences in SAC Index scores across estimated IQ groups. Levene's test revealed adequate homogeneity of variance in the Concentration and Delayed Recall total scores. SAC Orientation and Immediate Memory scores violated the homogeneity of variance assumption, thus the Welch statistic and Games-Howell test were used for primary and post-hoc analyses for those scores. Significant differences were observed among groups on the SAC Orientation total score, Welch's $F(2, 613.63) = 3.96$, $p < .05$. Post-hoc analyses revealed that individuals in the Above Average group ($M = 4.95$, $SD = 0.23$) performed significantly better than individuals in the Below Average group ($M = 4.88$, $SD = 0.36$, Cohen's $d = 0.23$). There were no significant differences between the Average group ($M = 4.91$, $SD = 0.31$) and either other group. Groups significantly differed in performance on the SAC Immediate Memory Index, Welch's $F(2, 608.59) = 24.90$, $p < .001$. Post-hoc analyses indicated that each group was significantly different from one another. Individuals in the Above Average ($M = 14.62$, $SD = 0.75$) group had significantly higher scores than individuals in both the Average ($M = 14.39$, $SD = 0.86$, Cohen's $d = 0.29$) and Below Average ($M = 14.15$ $SD = 1.02$, Cohen's $d = 0.53$) groups. Additionally, the Average group had significantly higher scores than the Below Average group (Cohen's *d*

Table 3.
Mean-level differences across WTAR standard score performance groups

Measure	Below Average	Average	Above Average	*F/Welch*	η^2
	Mean (SD)	Mean (SD)	Mean (SD)		
SAC Orientation	4.88 (0.36)	4.91 (0.31)	4.95 (0.23)	3.96*	0.01
SAC Immediate Memory	14.15 (1.02)	14.39 (0.86)	14.62 (0.75)	24.09**	0.03
SAC Concentration	2.85 (1.16)	3.34 (1.04)	3.86 (1.07)	72.69**	0.09
SAC Delayed Recall	4.32 (0.95)	4.38 (0.86)	4.44 (0.86)	1.50	0.00
ImPACT Visual Motor Speed Composite	36.74 (6.58)	39.55 (6.14)	42.74 (5.72)	55.46**	0.10
ImPACT Visual Memory Composite	75.90 (12.29)	77.28 (12.37)	82.08 (11.39)	19.71**	0.04
ImPACT Reaction Time Composite	0.59 (0.08)	0.57 (0.07)	0.55 (0.06)	18.86**	0.04
ImPACT Verbal Memory Composite	83.79 (9.65)	84.76 (10.29)	88.90 (10.13)	20.86**	0.03
Axon Processing Speed Score	105.23 (5.88)	105.51 (6.25)	106.67 (5.11)	4.77*	0.01
Axon Attention Score	106.10 (5.23)	107.10 (5.76)	108.22 (4.94)	7.28*	0.02
Axon Learning Score	98.89 (7.66)	99.14 (7.34)	100.67 (6.99)	4.37*	0.01
Axon Working Memory Speed Score	103.34 (6.60)	104.71 (6.51)	105.34 (6.32)	4.30*	0.01
ANAM Composite Score	0.06 (1.05)	0.49 (1.04)	0.93 (1.00)	34.80**	0.08
ANAM Simple Reaction Time Throughput Score	236.52 (28.99)	244.14 (25.95)	247.46 (24.61)	8.19**	0.02
ANAM Code Substitution Learning Throughput Score	57.57 (11.99)	59.89 (12.01)	63.57 (11.40)	13.54**	0.03
ANAM Procedural Reaction Time Score	101.03 (15.18)	104.70 (13.67)	108.98 (13.57)	16.16**	0.04

ANAM Math Processing Throughput Score	19.38 (6.23)	22.66 (6.32)	24.77 (6.21)	34.27**	0.08
ANAM Match to Sample Throughput	37.11 (12.68)	39.57 (12.11)	43.11 (12.71)	12.10**	0.03
ANAM Code Substitution Delayed Throughput Score	52.79 (13.76)	53.98 (14.96)	58.44 (15.18)	9.46**	0.02
ANAM Simple Reaction Time 2 Throughput Score	243.49 (35.23)	240.50 (30.05)	246.43 (25.97)	7.53*	0.02
ANAM Go/No-Go Difference Score	3.57 (1.39)	3.69 (1.39)	3.85 (1.37)	2.04	0.00

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

= 0.25). Significant group differences also emerged among SAC Concentration scores, *F(*2, 1403) = 72.69, $p < .001$. Post-hoc Bonferroni analyses indicated that each group significantly differed from one another, such that Above Average (*M* = 3.86, *SD* = 1.07) group outperformed both the Average (*M* = 3.34 *SD* = 1.04, Cohen's *d* = 0.48) and Below Average group (*M* = 2.85 *SD* = 1.16, Cohen's *d* = 0.89), while the Average group had significantly higher scores than the Below Average group (Cohen's *d* = 0.44). Finally, no significant differences between groups were noted on SAC Delayed Recall scores. For a complete summary of ANOVAs, see Table 3.

Table 4 reports actual rates of impaired scores obtained by participants across each measure. When considering observed SAC scores in the lowest 5% of the distribution, 12.40% of participants obtained at least one impaired score. Participants obtained at least two impaired scores at a rate of 1.00% and three impaired scores at a rate of 0.10%.

Table 4.
Actual frequency of impaired scores observed across measures

	Number of Impaired Scores Observed					
Measure	1	2	3	4	5	6
SAC	11.40%	0.90%	0.10	0.00%	0.00%	0.00%
ImPACT	10.50%	1.90%	0.10%	0.10%	0.00%	0.00%
Axon	12.70%	3.10%	1.10%	0.00%	0.00%	0.00%
ANAM	11.70%	3.30%	1.40%	0.80%	0.40%	0.10%

Note: Impaired scores were defined as occurring below the 5th percentile

Rates of impaired performance were also considered between groups. A Pearson Chi-square test for independence yielded a significant association between estimated IQ group an observed frequency of impaired scores on SAC tests, χ^2 (2, n = 1401) = 28.05, p < .001, Cramer's V = 0.14. Post-hoc tests revealed that individuals in the Above Average group obtained at least one impaired score at a significantly lower rate than expected. On the other hand, participants in the Below Average group yielded at least one impaired score at a significantly higher rate than expected. Whereas 19.90% in the Below Average group obtained at least one impaired score, the frequency was only 11.70% in the Average group and 8.10% in the Above Average group. Notably the rates of impairment observed in the below average group approximate those documented in Aim 1, whereas for individuals with average or greater intelligence, it was relatively uncommon to observe impaired scores. For a complete summary of Chi-square analyses, see Table 5.

Table 5.
Results of Chi-square analyses

Measure		Not Impaired	Impaired	Total	χ^2
SAC					28.05*
Below Average					
	Count	192	54	246	
	% Within Group	80.10%	19.90%	100.00%	

Adjusted Residual	-5.00**	5.00**		
Average				
Count	650	86	736	
%Within Group	88.30%	11.70%	100.00%	
Adjusted Residual	0.90	-0.90		
Above Average				
Count	385	34	419	
% Within Group	91.90%	8.10%	100.00%	
Adjusted Residual	3.20**	-3.20**		
ImPACT				23.80*
Below Average				
Count	137	27	164	
% Within Group	83.50%	16.50%	100.00%	
Adjusted Residual	-1.60	1.60		
Average				
Count	431	83	514	
% Within Group	83.9%	16.1%	100.0%	
Adjusted Residual	-3.30**	3.30**		
Above Average				
Count	287	15	302	
% Within Group	95.00%	5.00%	100.00%	
Adjusted Residual	4.90**	-4.90**		
Axon				9.58*
Below Average				
Count	109	27	136	
% Within Group	80.10%	19.90%	100.00%	
Adjusted Residual	-1.0	1.0		
Average				
Count	357	86	443	
% Within Group	80.6%	19.4%	100.0%	
Adjusted Residual	-2.1	2.1		
Above Average				
Count	230	28	258	
% Within Group	89.1%	10.9%	100.0%	
Adjusted Residual	3.1**	-3.1**		

ANAM				48.71*
Low Average				
Count	94	52	146	
% Within Group	64.4%	35.6%	100.0%	
Adjusted Residual	-6.2**	6.2**		
Average				
Count	363	77	440	
% Within Group	82.5%	17.5%	100.0%	
Adjusted Residual	0.2	-0.2		
Above Average				
Count	233	20	253	
% Within Group	92.1%	7.9%	100.0%	
Adjusted Residual	4.9**	-4.9**		

*Note: All analyses utilized Bonferroni corrections. * Denotes χ^2 values significant at the $p < .05$ level. ** denotes residuals significant at the $p < .008$ level.*

ImPACT

Aim 1. Bivariate Pearson correlations revealed significant correlations among all ImPACT composite scores. A moderate, positive correlation emerged between ImPACT Visual and Verbal Memory subtest scores ($r = 0.36$). A strong, negative correlation was observed between Simple Reaction Time and Visual Motor Speed ($r = -0.47$). While all subtests were significantly correlated with WTAR standard scores, the only relationship approaching moderate strength was between the WTAR standard score and ImPACT Visual Motor Speed ($r = 0.29$). For the complete correlation matrix, see Table 6.

Table 6.
Correlations among ImPACT scores

Measure	1.	2.	3.	4.	5.
N	399	399	399	399	396
1. Memory Composite Score (Verbal)	-				
2. Memory Composite Score (Visual)	.36**	-			
3. Visual Motor Speed Composite Score	.19**	.29**	-		
4. Reaction Time Composite Score	-.15**	-.18**	-.47**	-	
5. WTAR Standard Score	.13*	.11*	.29**	-.14**	-

Note: * indicates significance of $p < .05$, ** indicates significance of $p < .001$

Table 7 reports results of Monte Carlo analyses, utilizing different cutoffs to define impaired ImPACT scores. For example, when defining an abnormal score as occurring in the lowest 5th percentile of the normal distribution, 17.90% percent of the healthy population would be expected to obtain at least one impaired score. Using more liberal cutoffs for abnormality would increase this rate to 33.22% in the lowest 10th percentile and 46.49% in the lowest 15th percentile.

Table 7.
Simulated frequency of impaired scores on ImPACT

	Number of Impaired Scores Observed		
Cutoff	1	2	3
<5th Percentile	17.90%	1.98%	0.18%
<10th Percentile	33.22%	6.03%	0.75%
<15th Percentile	46.49%	11.55%	1.89%

Aim 2. One-way ANOVAs were completed to examine differences in ImPACT composite scores across estimated IQ groups. Levene's test revealed adequate homogeneity of variance among groups for all analyses. Groups significantly differed from one another on the Motor Speed Composite score, $F(2, 977) = 55.46$, $p < .001$. Post-

hoc Bonferroni analyses revealed the participants in the Below Average group had significantly lower ImPACT Motor Speed Composite scores (M = 36.74, SD = 6.58) than individuals in the Average (M = 39.55, SD = 6.14; Cohen's d = 0.44) or Above Average groups (M = 42.74 SD = 5.72; Cohen's d = 0.97). Performance on the Visual Memory Composite also differed by group, $F(2, 977) = 19.71$, $p < .001$. Individuals with in the Above Average group performed significantly better (M = 82.08, SD = .11.39) than the Average (M = 77.28, SD = 12.37; Cohen's d = 0.40) and Low Average groups (M = 75.90, SD = 12.29; Cohen's d = 0.52), which did not differ significantly from one another. On the ImPACT Verbal Memory Composite score, individuals in the High Average (M = 88.90, SD = 10.13) group performed significantly better than individuals in the Average (M = 84.76, SD = 10.29, Cohen's d = 0.41) or Low Average (M = 83.79, SD = 9.65, Cohen's d = 0.52) groups [$F(2, 977) = 20.83$, $p < .001$]. Similar to Visual Memory, there were no significant differences observed between the Low Average and Average group on ImPACT Verbal Memory. Finally, significant differences were observed among all groups on the ImPACT Reaction Time Composite score, $F(2, 977) = 18.86$, $p < .001$. Individuals in the Above Average group (M = 0.55, SD = .06) had significantly faster average reaction times than individuals in the Average (M = 0.57, SD = .07, Cohen's d = 0.31) or Below Average (M = 0.59, SD = .08, Cohen's d = 0.57) group. Additionally, individuals in the Average group had significantly faster reaction times than the Low Average group (Cohen's d = 0.27).

When considering the actual rate of performance below the 5th percentile, participants obtained at least one impaired score at a rate of 12.60%. While the majority of participants did not obtain an impaired score on this measure, as many as four impaired scores were observed. Over 2.00% of the sample obtained at least two impaired scores, and 0.20% obtained at least three, while another 0.10% obtained four impaired scores.

The actual rate of impaired performance on ImPACT tests was also considered among groups. A Pearson Chi-square test for independence yielded a significant association between estimated IQ group and rate of impaired score on ImPACT tests, χ^2 $(2, n = 980) = 23.80, p < .001$, Cramer's $V = 0.16$. Individuals in the Above Average group obtained at least one impaired score at a significantly lower rate than expected. In contrast, individuals in the Average group obtained at least one impaired score at a significantly higher frequency than anticipated. While 16.50% of individuals in the Below Average group and 16.10% of individuals in the Average group obtained at least one impaired score, the frequency was only 5.00% for the Above Average group.

Axon

Aim 1. Bivariate Pearson correlations indicated significant relationships among all Indices ($r \geq .14$), with the exception of the Learning and Processing Speed Indices. Additionally, strong positive correlations between Processing Speed and Attention subtests were observed ($r = 0.57$). Strong positive associations were also noted between Attention and Working Memory Speed subtests ($r = 0.52$), as well as a moderate positive association between Working Memory Speed and Processing Speed ($r = 0.40$). Axon subtests displayed only mild correlations with WTAR standard scores ($r \geq 0.11$), except the Processing Speed Index, which was not significantly correlated. For the complete correlation matrix, see Table 8.

Table 8.
Correlations among Axon scores

Measure	1.	2.	3.	4.	5.
N	377	377	377	377	376
1. Processing Speed Score	-				
2. Attention Score	.57**	-			
3. Learning Score	.03	.14**	-		
4. Working Memory Speed Score	.40*	.52**	.12*	-	
5. WTAR Standard Score	.05	.11*	.14**	.16**	-

Note: * indicates significance of $p < .05$, ** indicates significance of $p < .001$

Table 9 illustrates the results of Monte Carlo analyses at different cutoff points. These analyses revealed that when impaired scores are defined as those occurring below the 5th percentile, 16.01% percent of healthy individuals are expected to obtain at least one impaired score and 3.33% are expected to obtain at least two impaired scores. When expanding the cutoff to include the lowest 10% of the distribution, the frequency of observing an impaired score rose to 29.03% of healthy individuals obtaining at least one impaired score, and 8.51% of the distribution obtaining at least two impaired scores. At a cutoff of the lowest 15% of the distribution, about 40% of healthy individuals would be expected to obtain at least one impaired score, while more than 15% would be expected to obtain at least two impaired scores.

Table 9.
Simulated frequency of impaired scores on Axon

	Number of Impaired Scores Observed		
Cutoff	1	2	3
<5th Percentile	16.01%	3.33%	0.67%
<10th Percentile	29.03%	8.51%	2.26%
<15th Percentile	40.20%	14.56%	4.64%

Aim 2. One-way ANOVAs were computed to examine differences in performance on Axon subtests across estimated IQ groups. Levene's test revealed adequate homogeneity of variance among all groups except the Axon Processing Speed subtest. For that measure, the Welch statistic and Games-Howell test were used for primary and post-hoc analyses. Groups significantly differed in performance on the Axon Processing Speed subtest, Welch's $F(2, 266.52) = 4.77$, $p < .05$. Post-hoc Games-Howell tests revealed that individuals in the Above Average group ($M = 106.67$, $SD = 5.11$) performed significantly better on Axon Processing Speed subtests than individuals in either the Average ($M = 105.51$, $SD = 6.25$, Cohen's $d = 0.20$) or Below Average ($M = 105.23$, $SD = 5.88$, Cohen's $d = 0.26$) groups. There were no significant differences in performance between the Average and Below Average groups. Groups also significantly differed in performance on the Axon Attention subtest, $F(2, 845) = 7.28$, $p < .05$. Individuals in the Above Average ($M = 108.22$, $SD = 4.94$) group outperformed individuals in both the Average ($M = 107.10$, $SD = 5.76$, Cohen's $d = 0.21$) and Below Average ($M = 106.10$, $SD = 5.23$, Cohen's $d = 0.42$) groups. There were no significant differences in performance between the Average and Below Average groups. Additionally, group differences were observed in performance on the Axon Learning subtest, $F(2, 845) = 4.37$, $p < .05$. Individuals in the Above Average ($M = 100.67$, $SD = 6.99$) group

performed significantly better than individuals in the Average (*M* = 99.14, *SD* = 7.34, Cohen's *d* = 0.23) group. There were no significant group differences between the Low Average (*M* = 98.89, *SD* = 7.66) and either other group. Finally, there were significant differences in performance on the Axon Working Memory subtest. *F*(2, 845) = 4.30, $p <$ .05. On this measure, the Above Average (*M* = 105.34, *SD* = 6.32) group outperformed the Below Average (*M* = 103.34, *SD* = 6.60, Cohen's *d* = 0.31) group. There were no significant differences between the Average (*M* = 104.71, *SD* = 6.51) group and either of the other groups.

When considering the actual rate of obtaining scores below the 5th percentile, 16.90% of participants obtained at least one impaired scores on Axon measures. Participants obtained at least two impaired scores at a rate of 4.20% on this measure, and three impaired scores at a rate of 1.10%.

The actual frequency of obtaining one or more impaired scores was also examined by group. A Pearson Chi-square test for independence indicated a significant association between estimated IQ group and observed frequency of impaired scores on Axon tests, χ^2 (2, *n* = 837) = 9.58, $p < .05$, Cramer's *V* = 0.11. Participants in the Above Average estimate IQ group yielded one or more impaired scores at a significantly lower rate than expected. Participants in both of the other groups obtained impaired scores at a rate consistent with expectation. While 19.90% and 19.40% of individuals obtained at least one impaired score in the Below Average and Average groups, respectively, the frequency was only 10.90% for the Above Average group.

ANAM

Aim 1. Bivariate Pearson correlations revealed strong positive associations among many of the ANAM subtests ($r \geq 0.11$). A weak association was noted between the Go/No-Go difference score and the overall ANAM composite score ($r = 0.14$). As expected, subtests measuring the same construct (i.e. both Simple Reaction Time subtests [$r = 0.61$] and immediate and delayed Code Substitution Learning subtests [$r = 0.68$]) displayed strong positive correlations. Across ANAM subtests, weak to insignificant associations with WTAR standard score were noted ($r = 0.02 - 0.24$). For the complete correlation matrix, see Table 10.

Table 10.
Correlations among ANAM scores

Measure	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
1. Composite Score	-									
2. Simple Reaction Time Throughput Score	.56**	-								
3. Code Substitution Learning Throughput Score	.72**	.20**	-							
4. Procedural Reaction Time Throughput Score	.72**	.35**	.43**	-						
5. Math Processing Throughput Score	.56**	.11*	.33**	.35**	-					
6. Match To Sample Throughput Score	.68**	.23**	.45**	.39**	.28**	-				
7. Code Substitution Learning Delay Throughput Score	.61**	.11**	.68**	.29**	.21**	.37**	-			
8. Simple Reaction Time 2 Throughput Score	.67**	.61**	.28**	.42**	.20**	.30**	.21**	-		
9. Go/No-Go D Score	.14**	.02	.12*	.17**	.22**	.06	.02	.05	-	
10. WTAR Standard Score	.24**	.02	.22**	.20**	.22**	.15**	.19**	.11**	.06	-

Note: * indicates significance of $p < .05$, ** indicates significance of $p < .00$.

Table 11 displays the results of Monte Carlo analyses utilizing various cutoffs. These analyses revealed that when impaired scores are defined as falling below the lowest 5th percentile, at least 26.94% of the distribution would be expected to obtain at least one erroneously impaired score, while 9.72% would be expected to obtain at least two impaired scores and 4.70% would be expected to obtain at least three impaired scores. When the parameters for impaired scores are defined more liberally, the base rate of potentially erroneously impaired scores increases. Utilizing a cutoff of the lowest 10th percentile, the base rate of observing one impaired score rose to 44.89% and the base rate of observing two impaired scores rose to 21.76%. When defining impairment as the lowest 15th percentile, over half of the distribution would be expected to obtain at least one impaired score, and nearly a third would be expected to obtain at least two impaired scores.

Table 11.
Simulated frequency of impaired scores on ANAM

	Number of Impaired Scores Observed		
Cutoff	1	2	3
<5th Percentile	26.94%	9.72%	4.70%
<10th Percentile	44.89%	21.17%	11.53%
<15th Percentile	58.34%	32.47%	19.23%

Aim 2. A series of one-way ANOVAs were completed to examine mean-level differences in performances on ANAM subtests across estimated IQ groups. Levene's test revealed adequate homogeneity of variance among all groups except the ANAM Simple Reaction Time 2 subtest. For that measure, the Welch statistic and Games-Howell test were used for primary and post-hoc analyses. Significant differences were observed

between estimated IQ groups on ANAM composite scores, *F*(2, 836) = 34.80, *p* < .001. Bonferroni post-hoc analyses revealed significant differences among all groups. Individuals in the Above Average group (*M* = 0.93, *SD* = 100) outperformed individuals in both the Average (*M* = 0.49, *SD* = 1.04, Cohen's *d* = 0.43) and Below Average (*M* = 0.06, *SD* = 1.05, Cohen's *d* = 0.85) groups. Additionally, the Average group had significantly higher composite scores than the Low Average group (Cohen's *d* = 0.41). Groups significantly differed on the ANAM Simple Reaction Time Throughput score, *F*(2, 836) = 8.19, *p* < .001. Individuals in the Above Average group (*M* = 247.46, *SD* = 24.61) performed significantly better than individuals in the Below Average group (*M* =236.52, *SD* = 28.99, Cohen's *d* = 0.41). Additionally, the Average (*M* = 244.14, *SD* = 25.96) group significantly outperformed the Below Average group (Cohen's *d* = 0.31). There was no significant difference between the Above Average and Average group. Groups also differed in performance on the ANAM Code Substitution Learning Throughput score, *F*(2, 836) = 13.45, *p* < .001. Although there were no significant differences between the Below Average (*M* = 57.57, *SD* = 11.99) and Average (*M* = 58.89, *SD* = 12.01) groups, individuals in the Above Average (*M* = 59.89, *SD* = 12.01) group outperformed both the Average (Cohen's *d* = 0.31) and Below Average (Cohen's *d* = 0.51) groups. Additionally, groups differed in performance on the Procedural Reaction Time Throughput score, *F*(2,836) = 16.16, *p* < .001. Individuals in the Above Average (*M* = 108.98, *SD* = 13.57) group had significantly higher scores than individuals in either the Average (*M* = 104.70, *SD* = 13.67, Cohen's *d* = 0.31) or Below Average (*M* = 101.03, *SD* = 15.58, Cohen's *d* = 0.55) group. Additionally, the Average group significantly outperformed the Low Average group (Cohen's *d* = 0.25). Groups also significantly differed on ANAM Math Processing Throughput score *F*(2, 836) = 34.27, *p* < .001. On this measure, the Above Average (*M* = 24.77, *SD* = 6.21) group performed significantly better than both the Average (*M* = 22.66, *SD* = 6.32, Cohen's *d* = 0.34) and Below Average (*M* = 19.38, *SD* = 6.23, Cohen's *d* = 0.87) groups. Additionally, the Average group outperformed individuals in the Below Average group (Cohen's *d* = 0.52). Group differences were also observed on ANAM Match to Sample Throughput scores, *F*(2, 836) = 12.10, *p* < .001. The Above Average group (*M* = 43.11, *SD* = 12.71) performed significantly better than both the Average (*M* = 39.57, *SD* = 12.11 Cohen's *d* = 0.29) and Below Average (*M* = 37.11, *SD* = 12.68, Cohen's *d* = 0.47) groups. There was no significant difference in performance between the Average and Low Average groups. There were also significant differences among groups noted for ANAM Code Substitution Delayed Throughput scores, *F*(2, 836) = 9.46, *p* < .001. On this metric, the Above Average group (*M* = 58.44, *SD* = 15.18) outperformed both the Average (*M* = 53.98, *SD* = 14.96, Cohen's *d* = 0.30) and Below Average (*M* = 52.79, *SD* = 13.76, Cohen's *d* = 0.39) groups. There were no significant differences between the Below Average and Average groups. Significant group differences were also observed among groups on the second Simple Reaction Time Throughput score, Welch's *F*(2, 358.99) = 7.53, *p* < .05. Games-Howell post-hoc tests indicated that individuals in the Above Average (*M* = 246.43, *SD* = 25.97) group outperformed individuals in both the Average (*M* = 240.50, *SD* = 30.05, Cohen's *d* = 0.21) and Below Average (*M* = 243.49, *SD* = 35.23, Cohen's *d* = 0.09) groups. No significant difference in performance was observed between the Average and Below Average groups. Finally, there were no significant differences among groups on ANAM Go/No-Go Throughput scores.

Table 4 reports the actual frequency of scores below the 5th percentile obtained by participants. When examining actual rates of scores falling below the lowest 5th percent of the distribution, 17.70% of individuals obtained at least one impaired score on the ANAM test battery. While participants ranged from not obtaining any impaired scores to six impaired scores on this measure, 6.00% obtained at least two impaired scores, 0.80% at least three impaired scores, and 2.70% at least four impaired scores.

In considering group-level differences in impairment, a Pearson Chi-square test for independence indicated a significant association between estimated IQ group and rate of impaired score on ANAM tests, χ^2 (2, n = 839) = 48.71, $p < .001$, Cramer's V = 0.24. Individuals in the Above Average group obtained impaired at least one impaired score at a significantly lower rate than expected. In contrast, individuals in the Below Average group obtained one or more impaired scores at a significantly higher rate than would be expected. Whereas 35.60% of individuals in the Below Average group obtained at least one impaired score, the frequency was only 17.50% for the Average group and 7.90% for the Above Average group.

Discussion

Cognitive screening measures are routinely administered and interpreted to facilitate return-to-play decisions, despite their variable psychometric properties (Meehan et al., 2012). When interpreting test scores to make medical decisions, it is crucial to consider that healthy individuals regularly obtain "impaired" scores. These impaired scores may reflect "true" ability, testing error, fatigue, variable effort, and/or other factors inherent to neuropsychological testing. It is especially important to consider the rate at which individuals obtain impaired scores on CNT measures, as athletic trainers and coaching personnel rely on these measures to detect the acute symptoms of SRC. In the absence of any existing data on the rate at which healthy individuals obtain impaired scores on CNT measures, it is difficult to determine whether impaired scores reflect normal variability in performance or actual neurocognitive symptoms associated with SRC.

A primary aim of this research was to evaluate and compare expected base rates of impaired performance across four frequently utilized measures that purportedly detect cognitive symptoms associated with concussion. Ultimately, the test that yields the lowest expected rate of impaired scores would have the greatest clinical utility. As anticipated, Monte Carlo analyses revealed that it is likely that a significant percentage of non-injured athletes would obtain at least one impaired score on each measure investigated. Previous studies have suggested that gold-standard assessment instruments (i.e. the WAIS) are expected to yield erroneously impaired scores at a rate of about 20% (Crawford et al., 2007). The current study observed that rates varied somewhat across instrument and based on the criteria used to defined “impairment.” Overall, most measures yielded at least one impaired score (defined as falling below the 5th percentile) at a rate commensurate with previous studies. The SAC, ImPACT, and Axon yielded at least one impaired score at a rate of 17.68%, 17.90%, and 16.01%, respectively. The ANAM battery performed comparatively worse, yielding impaired scores at a much higher rate than other measures (26.94%). Noteworthy, Monte Carlo analyses suggest that it is relatively rare to observe two or more impaired scores in non-injured athletes (SAC 2.14%; ImPACT 1.98%; Axon 3.33%; ANAM 9.72%).

It was anticipated that CNT measures would consistently yield erroneous scores; however, inconsistent with hypotheses, the SAC did not appear to yield impaired scores a significantly lower rate than computerized measures. It was anticipated that CNT batteries would yield impaired scores at a higher rate than the SAC or traditional gold-standard assessment batteries (i.e. the WAIS) due to their variable psychometric properties (Randolph et al., 2005; Resch, et al., 2013).

Previous research does shed some light on the finding that approximately 20% of healthy athletes would be expected to obtain at least one impaired ANAM score. Rates of impaired scores increase as a factor of test battery size (Binder et al., 2009). Of all the CNT batteries evaluated, ANAM generated the most scores, increasing the likelihood of a healthy individual obtaining an impaired score. An additional factor to consider is that previous literature has also established that tests that are less directly associated with general intelligence are more likely to over- or under-estimate true ability (Donnell et al., 2011). While ANAM subtests exhibited only weak correlations with estimated IQ, it is unclear to what degree this might explain findings, as this was true for all SRC measures. The weak correlations observed with IQ (defined by reading ability) is not surprising given that, as a whole CNT, batteries tend to heavily sample domains of processing speed and sustained attention, which are only modestly associated with general intellectual ability.

In addition to documenting expected rates of impaired performance via Monte Carlo simulation, the current study aimed to explore whether rates of actual impaired performance differ among groups of individuals with different levels of intelligence. After stratifying the sample into groups based on estimated intellectual ability, actual rates of impaired scores were compared to those generated by Monte Carlo analyses. In general, across these three independent groups, participants obtained at least one impaired score at a lower rate than would be anticipated based on the Monte Carlo analyses and previous literature. While it was anticipated that about 20% of individuals would elicit at least one impaired score on each measure, the highest base rate of impaired scores was observed on the Axon at a rate of 12.70%. While base rates of impaired performance were expected to approach 20%, observed rates of impaired performance were closer to 10%. For example, while Monte Carlo simulations estimated that the ANAM battery would yield impaired scores at a higher rate than other measures (26.94%), participants actually obtained impaired scores on this measure at a rate of 11.70%. This rate was similar to the other measures, which yielded base rates from 10.50% to 12.70%.

A plausible explanation for the observed discrepancy in rates is the effect of intelligence as a moderator in CNT performance. In a 2015 critique of Odland and colleagues' (2011) application of Monte Carlo analyses to MMPI-2 scales, Tarescavage and Ben-Porath pointed out that that the Monte Carlo method assumes that all individuals have an equal likelihood of yielding an impaired score on a given measure. The authors argued that because neuropsychological tests are moderated by intellectual ability, there is not necessarily an equal likelihood of either outcome, which may cause the method to over- or underestimate the likelihood of obtaining an impaired score, particularly for individuals who have significantly higher- or lower-than-average intellectual ability

(Tarescavage & Ben-Porath, 2015). Similarly, Brooks and Iverson (2010) found that Monte Carlo analyses overestimated rates of impaired scores for individuals with higher-than-average intellectual ability while underestimating rates of impaired scores for individuals with lower-than-average intellectual ability. Collectively, this literature suggests that the current estimation of expected frequencies of impaired scores may be affected by the intellectual ability of participants.

It is well-documented that intellectual ability influences the likelihood of obtaining impaired scores. Consistent with Brooks and colleagues' (2009) research, the current study found that individuals with Above Average estimated IQ invariably obtained at least one impaired score at a rate that was significantly lower than anticipated. Conversely, on the SAC and ANAM, individuals with Below Average estimated IQ obtained impaired scores at a rate significantly higher than expected. Surprisingly, on ImPACT, individuals in the Average estimated IQ group yielded impaired scores at a higher rate than expected. Further, analyses of mean-level differences indicated that participants with higher IQ performed almost invariably better on CNTs, even if the measure was not significantly correlated with estimated IQ. Taken as a whole, it is clear that there are systematic differences in the way that individuals perform on CNT batteries based on intellectual functioning.

The current study has important implications for the clinical assessment and management of SRC. While graded symptom checklists remain the most commonly-used tool for the diagnosis and management of SRC, research has indicated that neurocognitive symptoms may continue to persist beyond the experience of physical symptoms of concussion (Harmon et al., 2013; Randolph, et al., 2005). As such, neurocognitive evaluation plays an important role in concussion management. As a result of increased awareness and availability of CNT batteries, a growing number of coaching and training personnel rely on these instruments to make clinical judgements. This is problematic as there is no strong body of evidence supporting that these measures are psychometrically sound. The current study illustrates, that like many other, traditional, neuropsychological instruments, popular CNT batteries may regularly elicit erroneously impaired scores. Neuropsychologists and other professionals trained in assessment may recognize this general limitation of neuropsychological assessment, and may carefully consider the meaning of impaired scores within the broader context of an evaluation. On the other hand, CNT batteries may be interpreted by a host of individuals with different training backgrounds, creating a higher likelihood that scores may be misinterpreted. The results of the current study suggest that impaired scores should be interpreted cautiously. Factors including the specific type of measure and the number of scores it generates should be taken into consideration when analyzing the results of a CNT instrument. While it is appreciated that not all athletic trainers are fully versed in the psychometric properties of the measures that they employ, this research suggests that they should carefully consider the meaning of an isolated impaired score, giving more credence to profiles generating multiple impaired scores.

Expanding upon ideas previously discussed regarding routine protocol for evaluation of concussion symptoms, it should be noted that many of these measures include symptom report scales that are routinely interpreted (e.g., ANAM Symptom Checklist, ImPACT Total Symptom Score). This research focused on exploring how the neurocognitive tasks function, but similar methods could be used to investigate the batteries more broadly. As an example, the SAC Total score is embedded within the SCAT3 (Guskiewicz et al., 2013), which also contains a symptom checklist (Symptom Severity Score) and test of balance (BESS). Monte Carlo analyses were conducted on these variables (see Tables 12 and 13). These analyses indicated that 14.22% of individuals would be expected to obtain at least one impaired score on the SCAT, while 0.76% would be expected to obtain at least two impaired scores and 0.01% would be expected to have three or more impaired scores. Compared to expected rates of impaired Index scores on the SAC, the SCAT yielded smaller rates of impairment. This discrepancy suggests that further research should be conducted to examine how symptom measures impact both the psychometric qualities and clinical utility of concussion measures.

Table 12.
Correlations among SCAT-2 scores

Measure	1.	2.	3.	4.
N	609	606	598	608
1. Symptom Severity Scale Score	-			
2. SAC Total Score	-.05	-		
3. BESS Total Score	.08*	-.14	-	
4. WTAR Standard Score	.01	.34**	-.09*	-

Note: * indicates significance of $p < .05$, ** indicates significance of $p < .001$

Table 13.
Simulated frequency of impaired scores on SCAT-2

	Number of Impaired Scores Observed		
Cutoff	1	2	3
<5th Percentile	14.22%	0.76%	0.01%
<10th Percentile	27.32%	2.62%	0.07%
<15th Percentile	38.99%	5.82%	0.26%

Beyond illustrating that CNT batteries routinely generate erroneously impaired scores, results of the current study also suggest that individuals of different intellectual abilities perform significantly differently on these measures, even when the tasks are not strongly related to general intelligence. Individuals with Above Average IQ not only consistently outperformed individuals of Below Average IQ, but also consistently obtained impaired scores at a significantly lower rate than expected. In sum, these results suggest that general intellectual ability plays an important role in how healthy individuals perform on CNT batteries, and that intellectual ability should routinely be assessed as part of baseline testing. Because intellectual ability meaningfully impacts test performance at baseline, knowledge of estimated IQ can inform the interpretation of CNT batteries, particularly when trying to understand them meaning of impaired scores. For example, individuals with lower intellectual ability may be more likely to obtain impaired scores at baseline, which may mask any change in scores during the acute post-concussion phase. Additionally, individuals with higher intellectual ability may have a lower chance of displaying impaired scores on CNTs, despite sustaining concussion. Currently, an evaluation of estimated intellectual ability is not a routine part of baseline testing; however, the results of this study suggest that even brief estimates of intellectual functioning may inform the interpretation of CNT batteries.

While the current study utilized one of the largest normative samples for each respective measure, it does not reflect the national demographic characteristics at large. These samples are primarily comprised of Midwestern athletes and have a greater proportion of males than females. Though males sustain concussion at a greater rate than females, in general, the results should be interpreted cautiously because there is evidence to suggest that female athletes may be more susceptible to sustaining a concussion during play. Additionally, their resulting recovery trajectory may differ from men (Broshek et al., 2011). In addition, the samples over-represent adolescents and young adults. The results of this study may not be readily generalizable to other populations, for example, older adults who also frequently sustain mTBIs. Future research should be conducted to explore the relationship among general intellectual ability and neurocognitive performance in more diverse and representative samples, including non-athletes.

Conclusions

The current study explored the psychometric properties of commonly-utilized CNT batteries to determine the rate at which healthy athletes obtain impaired scores at baseline. Consistent with previous research, the simulation suggested that about 20% of healthy athletes would be expected to obtain impaired scores across measures. When comparing actual performances across groups with different levels of general intellectual functioning, individuals with Above Average intellectual ability consistently outperformed individuals with lower intellectual ability, even when tasks were only modestly correlated with intelligence. Participants with Above Average intellectual ability almost invariably obtained impaired scores at a significantly lower rate than expected. In contrast, on some of the measures, individuals with Below Average intellectual ability obtained impaired scores at a significantly higher rate than expected. Given that athletic training and coaching personnel routinely use these instruments to evaluate cognitive status in the post-acute phase of concussion, impaired scores should be interpreted cautiously. These findings also suggest that it is important to incorporate measures of intellectual functioning in routine baseline cognitive assessments.

Bibliography

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.) Washington, D.C.: Author.

ANAM Sports Medicine Battery: Automated Neuropsychological Assessment Metrics (2010). Norman Center for the Study of Human Operator Performance.

Axelrod, B. N. & Wall, J. R. (2007). Expectancy of impaired neuropsychological test scores in a non-clinical sample. *International Journal of Neuroscience, 117*(11), 1591-1602.

Bailey, C., McCrea, M., & Barth J. (2013) Athletes and sport-related concussion: Complex issues in medical management. In Arcinigas, D.B., Zasler, N.D., Vanderploeg, R.D., & Jaffee, M.S. (Eds.) *Management of Adults with Traumatic Brain Injury.*

Barr, W. B. & McCrea M. (2001). Sensitivity and specificity of standardized neurocognitive testing immediately following sports concussion. *Journal of the International Neuropsychological Society, 7*(06), 693-702.

Barth, J. T., Freeman, J. R., Broshek, D. K., & Varney, R. N. (2001) Acceleration-deceleration sport-related concussion: the gravity of it all. *Journal of Athletic Training, 36*(3), 643-650.

Bazarian, J. J., Wong, T., Harris, M., Leahey, N., Mookerjee, S., & Dombovy, M. (1999). Epidemiology and predictors of post-concussive syndrome after minor head injury in an emergency population. *Brain Injury*, *13*(3), 173-189.

Binder, L. M., Iverson, G. L., & Brooks, B. L. (2009). To err is human: “Abnormal” neuropsychological scores and variability are common in healthy adults. *Archives of Clinical Neuropsychology*, *(24),* 31-46.

Broglio, S.P., Ferrara, M. S., Macciocchi, S. N., Baumgartner, T. A., & Elliott, R. (2007). Test-retest reliability of computerized concussion assessment programs. *Journal of Athletic Training, 42*(4), 509-14.

Brooks, B. L. & Iverson, G. L. (2010) Comparing actual to estimated base rates of “abnormal” scores on neuropsychological test batteries: Implications for interpretation. *Archives of Clinical Neuropsychology, 24*, 31-46.

Brooks, B. L., Strauss, E., Sherman, E. M. S., Iverson, G. L., & Slick, D. J. (2009). Developments in neuropsychological assessment: Refining psychometric and clinical interpretive methods. *Canadian Psychology/Psychologie Canadienne, 50*(3), 196-209.

Broshek, D. K., Kaushik, T., Freeman, J. R., Erlanger, D., Webbe, F., & Barth, J. T. (2005). Sex differences in outcome following sports-related concussion. *Journal of Neurosurgery*, *102*(5), 856-863.

Butcher, J. N. (2001). *Minnesota Multiphasic Personality Inventory-2: Manual for administration, scoring, and interpretation*. Minneapolis, MN: University of Minnesota Press.

CogState (2011). *CogState Research Manual (Version 6.0).* New Haven, CT: CogState Ltd.

Collie, A., Darby, D., & Maruff, P. (2001) Computerised cognitive assessment of athletes with sports related head injury. *British Journal of Sports Medicine, 35,* 297-302.

Crawford, J. R., Garthwaite, P.H., & Gault, C. B. (2007). Estimating the percentage of the population with abnormally low scores (or abnormally large score differences) on standardized neuropsychological test batteries: A generic method with applications. *Neuropsychology, 21*(4), 419-430.

Crawford, J. R., Garthwaite, P. H., Morrice, N., & Duff, K. (2012). Some supplementary methods for the analysis of the RBANS. *Psychological Assessment*, *24*(2), 365-374.

Crawford, J.R., Garthwaite, P.H., Sutherland, D., & Borland, D. (2011) Some supplementary methods for the analysis of the Delis Kaplan Executive Function System. *Psychological Assessment, 23(*4), 888-898.

Decker, S.L., Schneider, W.J., & Hale, J.B. (2012) Estimating base rates of impairment in neuropsychological test batteries: A comparison of quantitative models. *Archives of Clinical Neuropsychology, 27*, 69-84.

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

Donnell, A. J., Belanger, H. G., & Vanderploeg, R. D. (2011). Implications of psychometric measurement for neuropsychological interpretation. *The Clinical Neuropsychologist*, *25*(7), 1097-1118.

Echemendia, R. J., Putukian, M., Mackin, R. S., Julian, L., & Shoss, N. (2001). Neuropsychological test performance prior to and following sports-related mild traumatic brain injury. *Clinical Journal of Sport Medicine*, *11*(1), 23-31.

Ferguson, R. J., Mittenberg, W., Barone, D. F., & Schneider, B. (1999). Postconcussion syndrome following sports-related head injury: Expectation as etiology. *Neuropsychology, 13*(4), 582-589.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). “Mini-mental state”: a practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*(3), 189-198.

Gasonique, P. G. (1992). Affective state and awareness of sensory and cognitive effects after closed head injury. *Neuropsychology, 6*(3), 187-196.

Gessel, L. M., Fields, S. K., Collins, C. L., Dick, R. W., & Comstock, R. D. (2007). Concussions among United States high school and collegiate athletes. *Journal of Athletic Training*, *42*(4), 495.

Graham, R. Rivara, F. P., Ford, M. A., & Spicer, C. M. (2014). *Sports-Related Concussions in Youth: Improving the Science, Changing the Culture.* Washington, D.C.: The National Academies Press.

Green, P. (2004) *Medical Symptom Validity Test for Microsoft Windows User’s Manual*. Kelowana, BC: Green’s Publishing.

Guskiewicz, K. M., Mihalik, J. P., Shankar, V., Marshall, S.W., Bailes, J., Crowell, D. H., et al. (2007). Measurement of head impacts in collegiate football players; The relationship between head impact biomechanics and acute clinical outcome after concussion. *Neurosurgery, 61*(12), 1244-1253.

Guskiewicz, K. M., Register-Mihalik, J., McCrory, P., McCrea, M., Johnston, K., Makdissi, M., ... & Meeuwisse, W. (2013). Evidence-based approach to revising the SCAT2: introducing the SCAT3. *British Journal of Sports Medicine*, *47*(5), 289-293.

Harmon, K. G., Drezner, J., Gammons, M., Guskiewicz, K. M., Halstead, M., Herring, S….Roberts, W. (2013). American Medical Society for Sports Medicine position statement: Concussion in sport. *Clinical Journal of Sports Medicine, 23*(1), 1-18.

Harris, P. A., Taylor, R., Thielke, R., Payne, N., Gonzalez, N., & Conde, J. G. Research electronic data capture (REDCap) – A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Information*, *42*(2), 377-381.

Holm, L., Cassidy, J. D., Carroll, L. J., & Borg, J. (2005). Summary of the WHO collaborating centre for neurotrauma task force on mild traumatic brain injury. *Journal of Rehabilitation Medicine*, *37*(3), 137-141.

Hootman, J. M,, Dick, R., Agel, J. (2007) Epidemiology of collegiate injuries for 15 sports: summary and recommendations for injury prevention initiatives. *Journal of Athletic Training, 42*(2), 311-319.

Ingraham, L. J., & Aiken, C. B. (1996). An empirical approach to determining criteria for abnormality in test batteries with multiple measures. *Neuropsychology, 10*(1), 120-124.

Iverson, G. L. & Schatz, P. (2015). Advanced topics in neuropsychological assessment following sport-related concussion. *Brain Injury, 29*(2), 263-275.

Krauss, J. F., McArthur, D. L., Silverman, T., & Jayaraman, M. (1996). Epidemiology of brain injury. *Neurology and trauma. Houston: Saunders*, 3-17.

Lezak, M.D., Howieson, D.B., Bigler, E.D., & Tranel, D. (2012). *Neuropsychological Assessment* (5th ed.) New York, NY: Oxford University Press.

McCrea, M. (2008). *Mild Traumatic Brain Injury and Postconcussion Syndrome: The New Evidence Base for Diagnosis and Treatment.* New York, NY: Oxford University Press.

McCrea, M., Broshek, D. K., & Barth, J. T. (2015). Sports concussion assessment and management: Future research directions. *Brain Injury*, *29*(2), 276-282.

McCrea, M., Guskiewicz, K. M., Marshall, S. W., Barr, W., Randolph, C., Cantu, R. C., ... & Kelly, J. P. (2003). Acute effects and recovery time following concussion in collegiate football players: the NCAA Concussion Study. *JAMA, 290*(19), 2556-2563.

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

McCrea, M., Kelly, J., & Randolph, C. (2000). *Standardized Assessment of Concussion: Manual for administration, scoring, and interpretation* (2nd ed.) Waukesha, WI: CNS, Inc.

McCrory, P., Meeuwisse, W. H., Aubry, M., Cantu, B., Dvoøák, J., Echemendia, R. J., ... & Sills, A. (2013). SCAT3. *British Journal of Sports Medicine. 2013c*, *47*(5), 259-262.

Meehan, W. P., d'Hemecourt, P., Collins, C. L., Taylor, A. M., & Comstock, R. D. (2012). Computerized neurocognitive testing for the management of sport-related concussions. *Pediatrics*, *129*(1), 38-44.

Mild Traumatic Brain Injury Committee, American Congress of Rehabilitation Medicine, Head Injury Interdisciplinary Special Interest Group. (1993). Definition of mild traumatic brain injury. *Journal of Head Trauma and Rehabilitation, 8*(3), 86-87.

Nelson, L. D. (2015). False-positive rates of reliable change indices for concussion test batteries: A Monte Carlo simulation. *Journal of Athletic Training, 50*(12) 1319-1322.

Nelson, L. D., Janecek, J. K., & McCrea, M. A. (2013). Acute clinical recovery from sport-related concussion. *Neuropsychology Review, 23*(4), 285-299.

Nelson, L. D., LaRoche, A. A., Pfaller, A. Y., Lerner, E. B., Hammeke, T. A., Randolph, C., ... & McCrea, M. A. (2016a). Prospective, head-to-head study of three computerized neurocognitive assessment tools (CNTs): reliability and validity for the assessment of sport-related concussion. *Journal of the International Neuropsychological Society*, *22*(1), 24-37.

Nelson, L. D., Pfaller, A. Y., Rein, L. E., & McCrea, M. A. (2015). Rates and Predictors of Invalid Baseline Test Performance in High School and Collegiate Athletes for 3 Computerized Neurocognitive Tests ANAM, Axon Sports, and ImPACT. *The American Journal of Sports Medicine*, *43*(8).

Nelson, L. D., Tarima, S., LaRoche, A. A., Hammeke, T. A., Barr, W. B., Guskiewicz, K., ... & McCrea, M. A. (2016b). Preinjury somatization symptoms contribute to clinical recovery after sport-related concussion. *Neurology*, *86*(20), 1856-1863.

Odland, A. P., Martin, P., Perle, J., Simco, E. R., & Mittenberg, W. (2011). Rates of apparently abnormal MMPI-2 profiles in the normal population. *The Clinical Neuropsychologist*, *25*(7), 1134-1144.

Palmer, B. W., Boone, K. B., Lesser, I. M., & Wohl, M. A. (1998). Base rates of "Impaired" neuropsychological test performance among healthy older adults. *Archives of Clinical Neuropsychology, 13*(6), 503-511.

Rabinowitz, A.R., Li, X., & Levin, H.S. (2014). Sport and nonsport etiologies of mild traumatic brain injury: Similarities and differences. Annual Review of Psychology, 65, 301-331.

Randolph, C. (1998). *RBANS Manual.* San Antonio, TX: The Psychological Corporation.

Randolph, C., McCrea, M., & Barr, W. B. (2005). Is neuropsychological testing useful in the management of sport-related concussion? *Journal of Athletic Training, 40*(3), 139-152.

Resch, J. E., McCrea, M. A., & Cullum, C. M. (2013). Computerized neurocognitive testing in the management of sport-related concussion: An update. *Neuropsychology Review, 23*(4), 335-49.

Roberts, R.J., & Roberts, M.A. (2011) *Mild Traumatic Brain Injury: Episodic Symptoms and Treatment.* San Diego Plural Publishing.

Ruff, R. (2005). Two decades of advances in understanding of mild traumatic brain injury. *The Journal of Head Trauma Rehabilitation*, *20*(1), 5-18.

Ruff, R. M., Camenzuli, L., & Mueller, J. (1996). Miserable minority: emotional risk factors that influence the outcome of a mild traumatic brain injury. *Brain Injury*, *10*(8), 551-566.

Ruff, R. M., Iverson, G. L., Barth, J. T., Bush, S. S., Broshek, D. K., & NAN Policy and Planning Committee. (2009). Recommendations for diagnosing a mild traumatic brain injury: A national academy of neuropsychology education paper. *Archives of Clinical Neuropsychology : The Official Journal of the National Academy of Neuropsychologists, 24*(1), 3-10.

Schretlen, D. J., Testa, S., Winicki, J. M., Pearlson, G. D., & Gordon, B. (2008). Frequency and bases of abnormal performance by healthy adults on neuropsychological testing. *Journal of the International Neuropsychological Society, 14*(3), 436-445.

Sharpe, D. (2015). Your Chi-square test is statistically significant: Now what? *Practical Assessment, Research & Evaluation, 20*(8), 1-9.

Sheedy, J., Geffen, G., Donnelly, J., & Faux, S. (2006). Emergency department assessment of mild traumatic brain injury and prediction of post-concussion symptoms at one month post injury. *Journal of Clinical and Experimental Neuropsychology*, *28*(5), 755-772.

Short, P., Cernich, A., Wilken, J. A., & Kane, R. L. (2007). Initial construct validation of frequently employed ANAM measures through structural equation modeling. *Archives of Clinical Neuropsychology, 22, Supplement 1*(0), 63-77.

Sosin, D.M., Sniezek, J.E., & Thurman, D.J. (1996). Incidence of mild and moderate brain injury in the United States, 1991. *Brain Injury, 10*(1), 47-54.

SPSS Inc. (2016). SPSS base 24.0 for Windows user’s guide. SPSS Inc., Chicago, IL.

Summers, C. R., Ivins, B., & Schwab, K. A. (2009). Traumatic brain injury in the United States: an epidemiologic overview. *Mount Sinai Journal of Medicine: A Journal of Translational and Personalized Medicine*, *76*(2), 105-110.

Tarescavage & Ben-Porath, Y. S. (2015) A response to Odland et al.’s misleading,alarmist, estimates of risk for overpathologizing when interpreting the MMPI-2-RF. *The Clinical Neuropsychologist, 29*(2), 183-196.

Vanderploeg, R. D., Curtiss, G., & Belanger, H. G. (2005). Long-term neuropsychological outcomes following mild traumatic brain injury. *Journal of the International Neuropsychological Society*, *11*(03), 228-236.

Wechsler, D. (1997) *WAIS-III: Wechsler Adult Intelligence Scale, Third Edition*. San Antonio, TX: Pearson.

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

Wood, R. L., O'Hagan, G., Williams, C., McCabe, M., & Chadwick, N. (2014). Anxiety sensitivity and alexithymia as mediators of postconcussion syndrome following mild traumatic brain injury. *The Journal of Head Trauma Rehabilitation, 29*(1), 9-17.

Appendix A

Standardized Assessment of Concussion (SAC) Scores

Orientation

(Basic orientation, e.g. “What day of the week is it?”)

Neurological Screening*

(Basic neurological status, e.g. recollection of injury)

Concentration

(Digit span forward and backward)

Exertional Maneuvers*

(Performance of basic exercises during the memory delay period, e.g. jumping jacks)

Immediate Memory

(List learning initial trial)

Delayed Memory Recall

(Delayed list learning trial)

---

*subtest does not contribute to total score

Appendix B

IMPACT Scores

Verbal Memory Composite

Word Memory (Immediate and delayed word memory task)

Symbol Match (Visual learning and memory task with processing speed component)

Three Letters (Attention and working memory task with processing speed component)

Visual Memory Composite

Design Memory (Immediate and delayed visual memory task)

X's and O's (Sustained attention task)

Visual Motor Speed Composite

X's and O's (Sustained attention task)

Three Letters (Attention and working memory task with processing speed component)

Reaction Time Composite

X's and O's (Sustained attention task)

Symbol Match (Visual learning and memory task with processing speed component)

Color Match (Sustained attention task measuring response inhibition and reaction time)

Appendix C

AxonSports/CogState Measures

Detection

(Basic attention and processing speed task)

Identification

(Basic attention and processing speed task)

One Card Learning

(Visual learning and memory task)

One Back (Accuracy)

(Attention/working memory task, accuracy)

One Back (Reaction Time)

(Attention/working memory task, speed)

Two Back

(Attention/working memory task)

Groton Maze Learning Test

(Executive function and visuospatial learning task)

Groton Maze Learning Test—Delayed Recall

(Delayed visuospatial learning task)

Continuous Paired Associate Learning

(Nonverbal learning and memory task)

International Shopping List Task

(Verbal list-learning task, immediate recall)

International Shopping List Task—Delayed Recall

(Verbal list-learning task, delayed recall)

Social Emotional Cognition Test

(Tests ability to detect subtle changes in pictures)

Appendix D

ANAM Measures
Simple Reaction Time (Processing speed task)
Code Substitution Learning (Immediate, associative learning task)
Procedural Reaction Time (Attention and processing speed task)
Math Processing (Working memory task using basic arithmetic)
Match to Sample (Visual working memory task)
Code Substitution Delay (Delayed working memory task)
Simple Reaction Time 2 (Processing speed task)